U.S. Department of Justice
Office of Justice Programs
National Institute of Justice

NIJ

National Institute of Justice Discussion Paper

Predicting the Recidivism of Serious Juvenile Offenders

Richard L. Linster
Pamela K. Lattimore
Christy A. Visher

October 1990

National Institute of Justice
633 Indiana Avenue, NW
Washington, D.C. 20531
(202) 307-0144

128522

PREDICTING THE RECIDIVISM OF SERIOUS JUVENILE OFFENDERS
A DISCUSSION PAPER

Richard L. Linster
Pamela K. Lattimore
Christy A. Visher

National Institute of Justice
633 Indiana Avenue NW
Washington, DC 20531
(202) 307-0144

Opinions expressed in this paper are those of the authors and do not necessarily represent the positions of the U.S. Department of Justice or of the National Institute of Justice.

ABSTRACT

The latest report from the Federal Bureau of Investigation reveals that total arrests of persons under age 18 and arrests for index offenses among this groups increased 5.9 and 5.2 percent, respectively, between 1984 and 1988. Juvenile arrests from violent crime alone rose 7.7 percent in 1988. However, research on serious juvenile offenders has produced few results that focus on the practical aspects of reducing or controlling criminal activity in this population. In particular, little attention has been paid to developing and implementing techniques for assessing the risk of recidivism among adjudicated juvenile offenders.

This study seeks to fill some of the gaps in our knowledge about the serious juvenile offender and has several ogbjectives: (1) to examine recidivism patterns among a sample of youth released from the California Youth Authority; (2) to attempt to explain these patterns using a wide range of social characteristics and offending history variables; and (3) to develop preliminary risk assessment profiles for this sample based on statistical models.

The primary analytical task is to assemble the information contained in the data into an "intelligible" score function that can be demonstrated to have some credibility as an assessment of a subject's rearrest risk. We also discuss the practical application of these risk assessment profiles for decision making in a parole agency and for caseload management by individual parole officers.

TABLE OF CONTENTS

CHAPTER I
INTRODUCTION

1.1 Background

In recent years studies of serious youthful crime have produced few results that focus on the practical aspects of reducing or controlling criminal activity in this population. The lack of research on serious juvenile offenders is, in part, a consequence of the program of deinstitutionalization, diversion, and prevention outlined in the Juvenile Justice and Delinquency Prevention Act of 1974. This Act shaped the research agenda of the Federal Office of Juvenile Justice and Delinquency Prevention (OJJDP), the agency that supports most Federal research on juvenile offenders, for the next decade.

By the early 1980s, however, national statistics on juvenile crime indicated that more juveniles were being formally referred to court and that their offenses were more serious (Krisberg et al., 1986). And despite earlier predictions that juvenile crime would decline, juvenile arrests remained nearly constant between 1978 and 1982 (Cook and Laub, 1986). Moreover, the public was becoming increasingly concerned about juvenile crime and many in the research community believed that rehabilitation programs for juveniles did not work (Sechrest et al., 1979; Wright and Dixon, 1977). Thus, in 1984 a national committee recommended that the federal effort in the area of juvenile delinquency be redirected toward the control of the serious, violent, or chronic offender (see NAC, 1984; Regnery, 1986).

This new line of research is beginning to show some results. Although it is widely recognized that most juvenile offenders do not commit serious crimes repeatedly and do not commit crimes as adults (see Blumstein et al., 1986), a small group of youthful offenders (less than one-third) appear to be responsible for about 60% of all juvenile offenses (Bureau of Justice Statistics, 1988). Moreover, there is some evidence that serious juvenile crime is increasing. Total arrests of persons under age 18 and arrests for index offenses among this group increased 5.9% and 5.2%, respectively between 1984 and 1988 (FBI, 1989: Table 29). Juvenile arrests for violent crime alone rose 7.7% in 1988 (FBI, 1989: Table 31).

The juvenile justice system, however, does not appear well equipped to handle such increases (Cronin et al., 1988). Juvenile justice experts agree that the existing correctional options for serious persistent juvenile offenders are inadequate (Bishop et al., 1989; Krisberg et al., 1986; Speirs, 1988). Typical juvenile probation is generally considered to be ineffective because supervision is minimal and youth do not view probation as punishment.

The use of juvenile detention or other secure placement (e.g., "training schools") is rising (Krisberg et al., 1986) and severe crowding in juvenile institutions is occurring in some states (e.g., California). Given this dismal picture of juvenile justice, it is not surprising that recidivism among moderately serious, adjudicated juvenile offenders is high--67% in one 1982 study of a sample of over 3,000 male youth committed to probation camps in California who were followed for 24 months (Palmer and Wedge, 1989).

The challenge, then, is to develop methods for handling serious youthful offenders in the juvenile system. Several experts have suggested that the juvenile justice system and serious offenders, in particular, could benefit from some of the correctional innovations that have been introduced in the adult system in recent years (Baird et al., 1984). For example, risk assessment is widely used in the adult system, both for initial placement/classification and release.

An extensive literature exists on the use of offender classification for adults in the criminal justice system (see e.g., Gottfredson and Tonry, 1987; Farrington and Tarling, 1985). Although methodological issues are likely to be similar in adult and juvenile classification, substantive differences may well exist. Knowledge about the use of risk assessment and classification systems in juvenile justice decision making, especially for serious offenders, is limited. Moreover, those systems that are in use may be flawed because of their dominant concern with service needs and inattention to recidivism risk (Guarino-Ghezzi and Byrne, 1989).

1.2 A Portrait of the Serious Juvenile Offender

Much research and official data on juvenile criminal behavior shows that a small proportion of youth commit the majority of juvenile crime. What are the characteristics of these serious, persistent juvenile offenders? In a review of risk assessment instruments developed for use with juveniles in several states, Baird et al. (1984) identified eight factors associated with continued criminal involvement for juveniles: age at first adjudication, frequency and severity of prior criminal behavior, prior institutional commitments, alcohol and drug abuse, poor family relationships, negative peer influences, and school problems. However, no follow-up data or tests of the predictive accuracy of these factors is provided.

More sophisticated studies generally agree with this characterization of the influential life experiences of persistent juvenile offenders, albeit with some variations. These studies often incorporate more extensive information about the juvenile's family which allows further specification of the relevant family characteristics. These family influences include criminal father or siblings, poor parenting often involving ineffective supervision, and family conflict or disruption in family structure (see Blumstein et al., 1986; Greenwood, 1986; Loeber and Stoutheimer-Loeber, 1986; Elliott et al., 1985). These studies also find that a deprived background (low social class, poor housing, large family size) is characteristic of serious juvenile offenders. Additional factors related to chronic juvenile offending include poor school performance, early antisocial behavior (lying, stealing, "acting out"), and prior victimization.

One recent large study of young parolees (aged 17 to 22) gives some details about recidivism among a population that is closely related to the sample we examine in this paper. The report describes the criminal activities over a 6-year period of young offenders paroled in 1978 from prison in 22 states (Bureau of Justice Statistics, 1987). The study found that, overall, 69% had been rearrested, but that recidivism rates were highest in the first two years: 32% rearrested within one year and 47% within two years. Moreover, the length of the parolee's prior record was related to when the rearrest occurred: those with 4 prior arrests were twice as likely to be rearrested within the first year as those with 1 or 2 prior arrests.

Higher rates of recidivism were also found for young parolees who were incarcerated for a property offense, had a prior arrest for at least one violent offense, had been younger than 17 when first arrested as an adult, had not completed high school, or were under age 19 when paroled. Time served in prison, however, was not related to the likelihood of rearrest after parole.

1.3 Overview of Study

In a study of persistent juvenile offenders, recidivism is of course a prominent characteristic. Analyses of the 1945 and 1958 Philadelphia birth cohorts show that after three offenses (measured by police contacts), the probability of committing a fourth is about 0.72 and the recidivism probability is quite stable for subsequent offenses (see Weiner, 1989: Table 2.12). Thus, recidivism is likely to be high for juvenile offenders who engage in more than a few delinquent acts.

More useful to policymakers than whether or not a youth is likely to commit another offense would be knowledge about offense-specific patterns of recidivism, time to recidivism, and multivariate models which might predict various recidivist types. Maltz (1984) argues that failure-rate measures of recidivism (i.e., the time to failure) provide more information than typical recidivism measures such as the proportion of offenders who are rearrested (or reincarcerated) within some fixed time period. In fact, Maltz shows how the standard 1-year recidivism rate can produce misleading results in evaluations of the effectiveness of correctional programs in reducing recidivism. However, few studies of serious juvenile offending have examined juvenile recidivism using time-to-failure as the outcome measure of interest.

This paper seeks to fill some of the gaps in our knowledge about the serious juvenile offender by examining recidivism patterns among a sample of juvenile offenders released from California Youth Authority institutions in 1981-82. Juveniles in this particular sample had extensive criminal histories, especially for violent offenses, and the majority began their offending careers in their early teens.

Specifically, we use a multivariate survival model in an attempt to explain recidivism in this sample using a wide range of social characteristics and offending history variables. The study tries to distinguish among offenders based on the predicted risk of any rearrest within a specified time period (e.g. first three years after release).

We also discuss the practical applications of these preliminary risk assessment profiles for this sample from the perspective of a parole agency. For example, six-month-ahead "forecasts" of individual recidivism risk based on the statistical models might be used to distribute personnel and other resources among the current paroled population.

The paper is organized as follows: **Chapter II** discusses the nature of the sample and the variables used in the analyses, and presents some basic descriptive information on the sample subjects. **Chapter III** briefly introduces the specific form of the survival model used in the analyses and offers some tests of goodness of fit. (Further detail is given in the Appendices.) **Chapter IV** presents an interpretation of the results, particularly the effects of the socioeconomic and criminal history variables on recidivism. **Chapter V** discusses the predictive efficiency of the statistical model and its potential use as a classification instrument, using the model-derived forecasts of recidivism risk for selected offenders in the sample. **Chapter VI** considers the practical applications of the results for decision making in a parole agency and for caseload management by individual parole officers.

CHAPTER II
THE CYA RELEASE COHORT DATA

2.1 Overview of Chapter

The subjects of the current study are the 1949 male members of a randomly selected cohort of youths released to parole by the California Department of Youth Authority (or California Youth Authority, CYA) between July 1, 1981 and June 30, 1982.[1] The data were originally gathered by the CYA and the National Council on Crime and Delinquency (NCCD).[2] These data describe the subjects' criminal histories; instant commitments; personal and family characteristics; and arrests, convictions, and placements following release. As followup data were collected from California records, failure indicators (e.g., subsequent arrest) are for the State of California only. Followup data were available for at least three years for all subjects. The original data were augmented for the current study by adding county-level crime and clearance rates.

This chapter describes the data used in the analyses presented in subsequent chapters. The descriptive statistics convey some sense of the study population from the point of view of variables of interest to criminological theory. The following section examines the subject characteristics that comprise the dependent variable, time to failure, and the list of explanatory variables. As the correlations between the independent variables play an important role in the analyses presented in Chapter IV, these correlations are discussed in section 2.3. The characteristics of two sub-populations--early failures and long-term survivors--are then compared.

2.2 Subject Characteristics

The dependent variable for the current research is TIME, the length of time following arrest until "failure" -- defined here as first arrest or parole revocation.[3] Eighty-eight percent of the sample (1710 of 1949 subjects) "failed" during the followup period. For the subjects who failed, the mean time to failure was 306 days (standard deviation 293 days); the median was 204.5 days; and the modal failure time was 43 days (14 subjects). For the subjects who did not fail, the value of TIME was the length

---

[1]The original sample included 2200 males and females; the original investigators discarded 114 cases because of missing information from one or more sources. Of the 2086 remaining cases, 1998 were males. Forty-nine of these cases were dropped from the current study because of missing information for one or more variables. Note that this sample is __not__ a random sample of all California delinquents as criteria for committing delinquents to the CYA may vary between jurisdictions. Generally, only the most serious offenders are committed to the CYA.

[2]Funding and support for the collection of these data were provided by the David and Lucille Packard Foundation, the California Youth Authority, the National Council on Crime and Delinquency, and the Florence Burden Foundation. A report on the earlier study is provided in "Classification for Risk: The Development of Risk Prediction Scales for the Youth Offender Parole Board," Department of Youth Authority, Sacramento, CA, and the National Council on Crime and Delinquency, November 1987.

[3]For convenience, failure will subsequently be discussed as "first rearrest." Of the 1710 failures, 234 subjects (13.7 percent) failed due to revocation of parole.

of the followup period determined as the number of days between release and September 30, 1985.[4] The length of time to failure or censoring ranged from 1 to 1619 days.

Although the present study focuses on a single failure mode (i.e., any arrest or parole revocation), it is of interest to consider briefly the specific charge related to that failure. We identified five crime-specific failure modes. These were an arrest for (1) a violent offense, (2) robbery, (3) burglary, (4) other serious property offense, or (5) the ubiquitous "other" which for our subjects included misdemeanors and a variety of minor offenses.[5] This final category was considered to be representative of general delinquency.

The propensity towards violence is quite large in this cohort. With the exception of general delinquency (27.9 percent of failures), the most common reason for failure was a violent offense. Nearly 25 percent (24.3) of the cohort's first rearrests following release were for violent crimes. Combining robbery (9.4 percent of failures) with other violent offenses, about one-third of the cohort's initial rearrests included charges of violence. Burglary (18.8 percent) or other serious property crime (19.6 percent) was the most serious first rearrest charge for nearly 40 percent of the subjects.

A variety of socio-economic and criminal history variables that have been theoretically or empirically linked to offending are included in the analyses to be described in later chapters. The values of these variables provide a concise characterization of the CYA sample. Table II.1 lists these variables and their respective means and standard deviations.

2.2.1 Criminal History Variables

Focusing first on criminal history variables, it is apparent that as a group these youth began crime at an early age and have been fairly active. The average age of first arrest (AGEFIRST) was 14.2 years and the average time the subjects had engaged in crime (INCRIME), defined as the time between first arrest and the instant commitment, was 4.1 years. The subjects had been arrested an average of 7.58 times and more than 80 percent had four or more arrests (the value of NOARRSTS ranged from 1 to 30). About two-thirds (63.93 percent) had previously been committed to municipal, county, or state custody for a stay of more than 10 days; the mean number of previous commitments (PRCOMMIT) was 1.17. Most of the subjects (1033) had not previously violated parole, although numerous parole violations by some subjects resulted in a mean number of previous parole violations of 1.03 (14.4 percent had 3 to 14 previous parole violations).

---

[4]The records were searched during the last few months of 1985. September 30 was selected on the advice of CYA researchers as the latest date having reasonable assurance that all post-release records would be complete. First rearrests after September 30 were recorded for only four subjects.

[5]Violent offenses included homicide, assault, rape, weapons, and kidnapping. Robbery and burglary included these offenses as well as attempts. Serious property offenses included grand theft, auto theft, possession and sale of drugs, and arson. Other offenses, which were classified as "general delinquency," included miscellaneous assault (e.g., child endangering, riot, false imprisonment), petty theft, receiving stolen property, statutory rape, contributing to the delinquency of a minor, under the influence of drugs or alcohol, escape, miscellaneous felonies or misdemeanors, and welfare and institutional offenses. Up to three offenses per arrest were recorded for each subject. The most serious charge was used in determining the reason for failure.

TABLE II.1. SUBJECT CHARACTERISTICS[1]

Variable	Description	Mean/(SD)
Criminal History:		
AGEFIRST	Age at first arrest (years)	14.19/(2.81)
INCRIME	Time between first arrest & instant commitment (years)	4.14/(2.56)
NOARRSTS	Number of previous arrests	7.58/(4.64)
PPARVIOL	Previous parole violations	1.03/(1.43)
PRCOMMIT	Previous commitments (number > 10 days)	1.17/(1.20)
Scores:		
VIOLENCE	Violent criminal history score	1.22/(1.43)
ROBBERY	Robbery criminal history score	0.58/(0.87)
BURGLARY	Burglary criminal history score	1.66/(1.71)
OTHPROP	Serious property offense score	1.33/(1.56)
GENDELQ	General delinquency offense score	3.24/(2.83)
Current Commitment:		
MWF	Offense type (1 misdemeanor, 2 "wobbler", 3 felony)	2.39/(0.52)
CYAVIOL	Aggressive acts/threats during commitment (0 none,1 minor act <u>or</u> threat, 2 minor act <u>and</u> threat, 3 major act <u>or</u> threat, 4 major act <u>and</u> threat)	0.85/(1.30)
INFRRATE	Infraction rate (#infractions/timein)	0.82/(1.18)
TIMEIN	Length of confinement (years)	1.13/(0.61)
AGEOUT	Age at release (years)	19.45/(1.84)
Substance Abuse and School Problems:		
ALCOHOL	Alcohol abuse (0 if none, 1 if minor, 2 if major)	0.84/(0.81)
DRUGS	Drug abuse (0 if none, 1 if minor, 2 if major)	1.02/(0.80)
GANG	Gang involvement (0 if none, 1 if minor, 2 if major)	0.47/(0.79)
DROPOUT	School dropout (0 if no, 1 if yes)	0.55/(0.50)
SCHDISC	School discipline problems (0 if none,1 if minor, 2 if major)	0.81/(0.82)
Family Background:		
FAMSIZE	Number of siblings (0 if < 4, 1 if >= 4)	0.48/(0.50)
FAMVIOL	Intra-family violence or abuse (0 none; 1 minor violence <u>or</u> abuse; 2 major violence <u>or</u> abuse; 3 major violence <u>and</u> abuse)	0.40/(0.79)
PARALCH	Parental alc/drug dep. (0 none, 1 minor, 2 major)	0.46/(0.80)
PARCRIM	Parental criminality (0 none, 1 minor, 2 major)	0.32/(0.68)
SIBCRIM	Sibling criminality (0 none, 1 minor, 2 major)	0.65/(0.86)
WEAKMOM	Parental neglect/poor supervision (0 none; 1 minor neglect <u>or</u> supervision; 2 major neglect <u>or</u> supervision; 3 major neglect <u>and</u> supervision)	1.04/(1.00)
Environmental (subject's county of commitment):		
PCLRATE	Property crime clearance rate	0.17/(0.03)
PCRATE	Property crime rate	65.24/(10.66)
VCLRATE	Violent crime clearance rate	0.51/(0.09)
VCRATE	Violent crime rate	14.95/(4.49)

1 N = 1949.

In an effort to capture both the nature and extent of past criminal activity, we developed individual criminal history score variables that condensed all previous arrest charges into the five "Scores" variables shown in Table II.1. (In the original data, as many as three charges were included for each arrest.) Scores were developed for the following five categories: VIOLENCE, ROBBERY, BURGLARY, OTHPROP (other property), and GENDELQ (general delinquency). (The offenses which comprise these categories are discussed in footnote 5.)

Each of the variables is a weighted sum of the number of charges for offenses of each of the five types.[6] The weight is a function of the number of days between the offense and the instant commitment to CYA. All offenses that occurred within two years (730 days) of the instant commitment are given a weight of 1; earlier offenses are downweighted by the ratio 730/(number of days from arrest on that charge to instant CYA commitment). Thus, for example, a charge three or four years earlier is given a weight of 0.667 or 0.5, respectively.

The rationale for the weighting scheme is that offenses that occurred more than 2 years earlier might reasonably be assumed to have less bearing on current criminal propensities than those that occurred more recently. Not surprisingly, given the age of the subjects, the highest mean score (2.19) occurred in the general delinquency (GENDELQ) category; values on this score ranged from 0 to 19.6. Considerable past involvement in more serious offending was also indicated by the mean scores for BURGLARY (1.66), OTHPROP (1.33), and VIOLENCE (1.22).

These values indicate that, on average, members of this cohort had more than one previous arrest with a charge for crimes within each of these categories. Maximum values on these variables are also informative of the extent to which these youth engaged in crime. The maximum values of BURGLARY, OTHPROP, and VIOLENCE were 11.6, 13.7, and 10.1, respectively. The smallest average value occurred in the ROBBERY score, with the average equal to 0.58 (the maximum value on the robbery score was 6.5).

2.2.2 Current Commitment

Information on the instant offense and the behavior of the individuals while committed for the instant offense was also available. The variable MWF refers to the (most serious) charge that led to the instant CYA commitment. An MWF value of "1" indicates that the offense was a misdemeanor and a value of "3" indicates a felony. The intermediate score of "2" is defined as a "wobbler." In other words the offense can be either a misdemeanor or a felony. This variable, thus, provides a measure of the seriousness of the instant offense. The mean value of MWF was 2.39; the modal value was 2 (57.7 percent). Less than 2 percent (1.74 percent) of the cohort's most serious commitment offenses were classified as misdemeanors.

Two variables were created as measures of the behavior of the subjects during their confinement in CYA facilities. The first, INFRRATE, measured the number of disciplinary infractions as a function of time in custody. The cohort averaged 0.82 infractions per year. The second variable, CYAVIOL, combined measures in the original data of overt aggressive behavior and threats made during confinement. The value of this variable ranged from 0 (no aggressive behavior or threats) to 4 (major evidence of both overt aggressive acts and threats). For most subjects (1214, 62.3 percent), the records indicated no aggressive behavior or threats (CYAVIOL = 0). The records of an additional 364 subjects (18.7 percent) suggested only minor evidence of aggressive behavior during confinement. The

---

[6]Any criminal history score function will necessarily be somewhat arbitrary. For a recent comparison of a variety of operational definitions of prior criminal history, see Nelson 1989.

mean of this indicator variable was 0.85. The average length of confinement (TIMEIN) was 1.13 years and the average age at release (AGEOUT) was 19.45.

2.2.3 Substance Abuse and School Problems

The data suggest that, on average, most of these subjects were not seriously involved in substance abuse or gangs prior to their instant CYA commitment. The means of the variables ALCOHOL, DRUGS, and GANG were 0.84, 1.02, and 0.47, respectively. (These indicator variables were scored by the original investigators as 0 if there was no evidence in the subject's records of the behavior, 1 if the evidence suggested minor involvement or problems, and 2 if the evidence suggested major involvement or problems.) Thirty-three percent of the sample had "major" problems with drugs, while 26 percent had "major" problems with alcohol. Twenty-nine percent of the cohort had some (either minor or major) association with a gang prior to confinement.

Fifty-five percent of the subjects had quit school (DROPOUT = 0.55) and slightly more than half (55.4 percent) had records suggesting school disciplinary problems. The mean of the variable SCHDISC was 0.81, where the coding of the variable was again 0, 1, and 2.

2.2.4 Family Background

Six variables provide information on the characteristics of the subjects' families. The subjects were as likely as not to have 4 or more siblings (the mean of the dichotomous variable FAMSIZE was 0.48). For 62.1 percent of the sample, there was evidence of some parental neglect or poor parental supervision. This variable, WEAKMOM, had an average value of 1.04 (where a value of 1 indicates minor evidence of neglect and poor supervision). The variable FAMVIOL provides a measure of both evidence of intra-family violence (any family members) and subject-specific abuse. This variable, which could have the value of 0, 1, 2, or 3, where higher values indicated more serious evidence of family violence, had a value of 0 (no evidence of violence or abuse) for 75.9 percent of the subjects. The mean value was 0.40.

The final three variables which characterize the subjects' families concern parental alcoholism (PARALCH) and evidence of parental and sibling criminality (PARCRIM and SIBCRIM). Most subjects' records indicated no evidence of parental alcohol problems (PARALCH = 0 for 73.1 percent) or parental or sibling involvement in crime (PARCRIM = 0 for 79.8 percent, SIBCRIM = 0 for 60.7 percent). The mean values for PARALCH, PARCRIM and SIBCRIM were 0.46, 0.32, and 0.65, respectively. (These variables were coded as 0 for no evidence, 1 for minor evidence, and 2 for major evidence.)

2.2.5 Environment

A final group of variables included in the analysis were two types of "environmental" variables. The first type was comprised of the county-level Uniform Crime Report crime and clearance rates for property and violent index offenses. Values for these variables were associated with each subject by his county of commitment. As can be seen in Table II.1, the mean property crime and crime clearance rates were 65.24 crimes/1000 population and 0.17 clearances/reported crime, respectively. The mean violent crime and crime clearance rates were 14.95/1000 population and 0.51 clearances/reported crime, respectively. Values of these variables varied considerably within the State of California, with, for example, the violent crime rate ranging from 4.82 to 26.42 and the violent crime clearance rate ranging from 0.19 to 0.82.

In order to capture the possible influence of gross differences in environment not measured by county-level crime and clearance rates, a second type of environmental variable was created that is composed of four regional variables (LA, BAYAREA, SONOTLA, and NORCNTRL).[7] Values for these variables (0 or 1) were associated with each subject by his county of commitment. The distribution of N over these four regional variables is shown below:

Region	Description	Number	Percent
LA	Los Angeles	769	0.395
BAYAREA	San Fran. Bay	394	0.202
SONOTLA	So. CA, not LA	365	0.187
NORCNTRL	North/Central	421	0.216

An examination of means of the regional sub-samples indicates that there are differences between regions in population characteristics. Table II.2 illustrates this with a few sample means for the populations from Los Angeles (LA) and from the North/Central (NORCNTRL) region. As can be seen, the LA sample contains a higher concentration of offenders with records of substantial involvement in violence, robbery and serious property crime. Plausible interpretations of the information include: (1) LA youth exhibit a higher propensity towards serious crime or, perhaps more likely, (2) the data reflect a regional difference in policy with regard to commitment to the CYA. If we may reasonably assume that the prevalence of the relatively less serious offenses grouped into GENDELQ is not significantly smaller in LA than elsewhere in the State, it would appear that such cases are more likely to be sent to the CYA if adjudicated in the North/Central region than in Los Angeles county.

2.3 Correlations

While each of the variables discussed above measures conceptually distinct theoretical constructs, they are obviously not orthogonal. Correlations between independent variables are of importance in determining the values assigned to the model's parameters but their role is hidden in the mathematics of solution of the likelihood maximization equations. They can, however, be used explicitly in a simple assessment of the relative importance of different variables in the model's assignment of risk--a subject to which we shall return in Chapter IV.

Tables II.3 and II.4 present selected correlations that are significant at the 0.05 level.[8] Table II.3 includes all of the criminal history and current commitment variables (see Table II.1). Three substance abuse/school problem variables (GANG, DROPOUT, and SCHDISC) and one family background variable (SIBCRIM) with correlations larger than |0.20| with one or more of the fifteen criminal history/current commitment variables are also included. Table II.4 includes the fifteen substance abuse/school problems/environment variables in Table II.1, as well as four criminal history variables (AGEFIRST, NOARRSTS, VIOLENCE, and AGEOUT) which had correlations greater than about |0.20| with one or more of the other fifteen variables. This division of the correlation matrix is

[7]LA is only Los Angeles County; BAYAREA is comprised of the counties of Sonoma, Napa, Solano, Marin, Contra Costa, Alameda, San Francisco, San Mateo, and Santa Clara; SONOTLA includes the Southern California Counties of Santa Barbara, Ventura, San Bernardino, Orange, Riverside, San Diego, and Imperial; NORCNTRL includes all other counties.

[8]In a sample of 1949 subjects, correlations between pairs of variables greater than about |0.045| are significantly different from zero at the 0.05 level (two-tailed test).

TABLE II.2. COMPARISON OF REGIONAL SAMPLE MEANS

Variable	LA Region	NORCNTRL Region
TIMEIN	1.22	1.00
VIOLENCE	1.28	1.07
ROBBERY	0.81	0.32
OTHPROP	1.49	1.19
GENDELQ	2.49	3.77
PRCOMMIT	0.95	1.34
AGEOUT	19.76	19.07
GANG	0.83	0.18
N	769	421

appropriate since, with few exceptions, the strongest correlations occurred within groups of variables: criminal history measures, drug and alcohol use, family problems, and crime and criminal justice environment variables.

For the most part, the discussion of correlations among the variables (Tables II.3 and II.4) will address only those variables with correlations greater than |0.20|. The rationale for this criterion, in addition to parsimony, is to restrict the discussion to only those correlations which are meaningful. The correlation matrices are discussed in the next two sections.[9]

2.3.1 Correlations among the Criminal History Variables

The first variable in Table II.3 is AGEFIRST, age at first arrest. AGEFIRST is negatively correlated with the length of time since first arrest, number of arrests, number of previous parole violations, number of previous CYA commitments, three criminal history scores (BURGLARY, OTHPROP, and GENDELQ), school disciplinary problems, and evidence of sibling criminality. The signs of these correlations are what one would expect on theoretical grounds (e.g., it is reasonable that those who were arrested at younger ages are more likely to have more prior commitments).

What is perhaps most striking is that the magnitudes are as small as they are. For example, while variation in age at first arrest would (in a simple linear model) explain about 40 percent of the variation in numbers of arrests, it would explain only about 9 percent of the variation in prior commitments. It is perhaps also noteworthy that age at first arrest is not statistically related to the robbery score. Only one variable is positively correlated with age at first arrest, AGEOUT. Given the negative association between age at first arrest and other measures of the extent of criminal his-

[9]Although a correlation of about 0.045 is statistically significant at accepted levels for a sample of 1949, this correlation suggests that only about 0.2 percent of the variance of one of the variables is "explained" by the other variable in the pair. Correlations of |0.20| and above suggest explanatory power of 4 percent or more.

tory, this positive association between ages at first arrest and at release could not have been anticipated. Despite generally longer criminal careers, subjects in this population whose first arrest occurred earlier still tend to be younger at the time of release from the instant commitment.

Five variables, numbers of arrests, parole violations and previous commitments, and the history scores BURGLARY and GENDELQ, were positively correlated with INCRIME, the length of time between first arrest and the instant commitment.[10] In a simple linear model relating number of arrests to career length, the slope would have a value of about 1.2 arrests/year active and variation in career length would explain about 44 percent of the variance in numbers of arrests. Correlations with other criminal history variables also have the expected sign but the explained variances in simple linear models are all rather weak.

Number of arrests (NOARRSTS) is, as would be expected, positively correlated with six criminal history variables (in addition to INCRIME). The strongest correlation is between the general delinquency score (GENDELQ) and number of arrests (r = 0.69). This relationship is an indication of the frequency with which such charges appear in subjects' arrest records. (The mean value of GENDELQ was 3.24 for these subjects; see Table II.1). The other positive correlates to number of arrest are previous parole violations, prior commitments, sibling criminality, and history scores for VIOLENCE, BURGLARY, OTHPROP. Number of arrests is strongly and negatively correlated with only one variable, AGEFIRST (as previously discussed).

The number of previous parole violations (PPARVIOL) is strongly correlated with only one variable other than those discussed previously, the general delinquency history score. The positive relationship between these two variables (r = 0.21) is in the expected direction. Similarly, the number of previous commitments (PRCOMMIT) is positively related to two of the criminal history scores in addition to CARLNGTH and NOARRSTS. These variables are OTHPROP and GENDELQ.

In addition to age at first arrest, only one variable is negatively associated with the number of previous commitments, MWF (by our 0.20 criterion). The negative correlation between PRCOMMIT and MWF, our measure of instant offense seriousness, suggests that those with more prior commitments are less likely to have been committed for a felony than a misdemeanor. This somewhat counter-intuitive result may be due to the high incidence of general delinquency among this population.

One or more of the criminal history scores, with the exception of ROBBERY, have been seen to be strongly correlated with the five variables discussed above. Looking first at the relationship between pairs of these history variables, only one correlation meets our 0.2 criterion, the negative association between ROBBERY and BURGLARY (r = -0.20). This association suggests that burglars are less likely to be robbers and vice versa. Two other negative correlations are statistically significant among these variables (although failing our 0.20 criterion)--VIOLENCE and BURGLARY (r = -0.16) and ROBBERY and GENDELQ (r = -0.15). These relationships suggest differences in individuals who engage in violent crimes (VIOLENCE or ROBBERY) and those who engage in property crimes or minor offenses.

[10]The correlations between career length and the violence history score and time in custody were statistically significant (r = -0.15 and -0.16, respectively), suggesting that shorter careers were associated with greater violent offending and longer sentences.

TABLE II.3. CORRELATION MATRIX: CRIMINALITY VARIABLES[1]

	AGEFIRST	INCRIMES	NOARRSTS	PPARVIOL	PRCOMMIT	VIOLENCE	ROBBERY	BURGLARY	OTHPROP	GENDELQ	MWF	CYAVIOL	INFRRATE	TIMEIN	AGEOUT	GANG	DROPOUT	SCHDISC	SIBCRIM
AGEFIRST	1	-.77	-.62	-.21	-.30	-.15		-.28	-.24	-.39	.16	-.07	-.09		.47	-.08	.17	-.22	-.20
INCRIME		1	.66	.25	.27	.11		.22	.18	.30	-.08		.06	-.16	.17			.06	.18
NOARRSTS			1	.27	.41	.23		.37	.37	.68	-.19		.10	-.14	-.08	.12	-.05	.13	.20
PPARVIOL				1	.13			.05	.06	.21	-.07		.06	-.16		.08		.08	
PRCOMMIT					1		-.07	.19	.21	.38	-.20		.08	-.12	-.12			.06	.08
VIOLENCE						1		-.16	-.06	.14		.17	.08	.19		.24		.14	.11
ROBBERY							1	-.20		-.15	.42	.09		.19		.12		.05	
BURGLARY								1	.06	.13	-.25	-.07		-.12	-.17	-.07	-.07		.10
OTHPROP									1	.15	-.15	-.05		-.06	-.12		-.06		
GENDELQ										1	-.29		.08	-.19	-.23	.05	-.09	.16	.12
MWF											1	.09		.36	.25		.07	-.05	
CYAVIOL												1	.45	.29		.14		.13	.05
INFRRATE													1	.22		.05		.07	
TIMEIN														1	.15	.08			-.05
AGEOUT															1	-.09	.30	-.23	-.07
GANG																1		.16	.08
DROPOUT																	1		
SCHDISC																		1	.08
SIBCRIM																			1

1 N = 1949. Only correlations greater than or equal to |0.045| are included. Correlations of |0.45| and larger are significantly different from zero at the $\alpha = 0.05$ level (two-tailed test). In addition to the 15 criminal history variables (AGEFIRST...AGEOUT), the four individual/family variables with correlations with the criminal justice variables larger than about 0.20 are also included in this table.

TABLE II.4. CORRELATION MATRIX: INDIVIDUAL/ENVIRONMENT VARIABLES[1]

	ALCOHOL	DRUGS	GANG	DROPOUT	SCHDISC	FAMSIZE	FAMVIOL	PARALCH	PARCRIM	SIBCRIM	WEAKMOM	PCLRATE	PCRATE	VCLRATE	VCRATE	AGEFIRST	NOARRSTS	VIOLENCE	AGEOUT
ALCOHOL	1	.39	.05	.13	.08		.08	.12			.13	.08	-.08	.09	-.16	.09			.12
DRUGS		1	.09	.10	.09		.09	.10	.05	.06	.16						.09	-.06	.13
GANG			1		.16	.08	-.05		.04	.08	.07	-.24		-.19	.13	-.08	.12	.24	-.09
DROPOUT				1		.06						-.09		-.06		.17	-.05		.30
SCHDISC					1		.07	.07	.13	.08	.23					-.22	.13	.14	-.23
FAMSIZE						1				.21			.06		.06			.10	.06
FAMVIOL							1	.33	.15		.24				-.06	-.04			-.08
PARALCH								1	.33		.20	.05			-.05	-.07	.06		-.07
PARCRIM									1	.08	.22					-.14	.10		-.15
SIBCRIM										1	.11				.06	-.20	.20	.11	-.07
WEAKMOM											1					-.19	.13		-.20
PCLRATE												1	-.16	.64	-.24				-.13
PCRATE													1	.10	.61				
VCLRATE														1	-.15		-.07		-.05
VCRATE															1				
AGEFIRST																1	-.62	-.15	.47
NOARRSTS																	1	.23	-.08
VIOLENCE																		1	
AGEOUT																			1

1 N = 1949. Only correlations greater than or equal to ¦0.045¦ are included. Correlations of ¦0.45¦ and larger are significantly different from zero at the α = 0.05 level (two-tailed test). In addition to the 15 individual/family/ enviromental variables (ALCOHOL...VCRATE), the four criminal history variables with correlations with the criminal justice variables larger than about 0.20 are also included in this table.

Relationships between criminal history scores and other variables are in the expected direction. High violence scores are associated with gang involvement; those with high ROBBERY scores are more likely and those with high BURGLARY and OTHPROP scores are less likely to have the instant offense be classified a felony (i.e., MWF = 3). Finally, those with high general delinquency scores are likely to be younger at release.

MWF, the measure of instant offense seriousness, is strongly associated with two variables, TIMEIN and AGEOUT, in addition to the relationships discussed earlier. Both of these associations are positive, as would be expected, since, loosely, they imply that a felony is associated with a longer time served and an older age at release.

Two variables measure the subjects' behaviors during their institutionalization, CYAVIOL and INFRRATE. Not surprisingly, those with the most evidence of violent behavior and the higher rates of infractions are most likely to have longer commitments (TIMEIN).

The length of the instant commitment (TIMEIN) is, as would be expected, strongly (and positively) related with the nature of the instant offense (MWF) and conduct during the commitment. TIMEIN is also positively correlated with the scores for VIOLENCE and ROBBERY (r = 0.19 for both). Curiously, however, TIMEIN is negatively correlated with three simple measures of extent of criminal history: length of time since first arrest, number of previous arrests, and previous parole violations (r = -0.16, -0.14, and -0.16 for INCRIME, NOARRSTS, and PPARVIOL, respectively). These results, again, may be related to the prevalence of minor offending in this population (i.e., high GENDELQ scores) since TIMEIN is also negatively correlated with this criminal history score (r = -0.19). TIMEIN is positively correlated with age at release (AGEOUT, r = 0.15). AGEOUT is also strongly correlated with the two measures of educational behavior, DROPOUT and SCHDISC. Those older at release are more likely to have dropped out of and to have experienced school disciplinary problems while in school.

2.3.2 Correlations among the Substance Abuse/School/Family/Environment Variables

The first five variables in Table II.4 refer to individual characteristics--the indicators of substance abuse (ALCOHOL and DRUGS), gang involvement (GANG), and the two "school" variables DROPOUT and SCHDISC. One of the strongest correlations among individual characteristics is the positive correlation between alcohol and drug abuse (r = 0.39). Strong associations between the variable GANG and other variables were found in only two cases, property crime clearance rate, PCLRATE, (r = -0.24) and, as previously discussed, the VIOLENCE history score (r = 0.24). Both DROPOUT and SCHDISC were strongly associated with age at release (AGEOUT)--DROPOUT positively and SCHDISC negatively. It is not surprising that older subjects would have been more likely to quit school, but the negative correlation with SCHDISC--implying younger subjects were more likely to have had disciplinary problems in school--is not as self evident. SCHDISC was also negatively correlated with AGEFIRST, implying that those with more disciplinary problems were more likely to have begun their criminal careers at a younger age. Finally, parental neglect and poor parental supervision (WEAKMOM) were positively correlated with SCHDISC.

The next six variables are family measures. Overall, the correlations suggest a consistent tendency to indicate a syndrome of family pathology. The 0.32 correlation between FAMSIZE and SIBCRIM, while strong, may simply be a matter of variable definition. Family size itself may or may not have an effect of criminal behavior, but certainly, the larger the number of siblings, the greater the chance that at least one of them will have a criminal record. Evidence of family violence (FAMVIOL) is strongly and positively correlated with both parental alcohol abuse (PARALCH) and poor parenting (WEAKMOM). PARALCH, in turn, is associated with a greater likelihood of parental criminality (PARCRIM) and poor parenting (WEAKMOM). WEAKMOM is positively correlated with the number of arrests (NOARRSTS) and evidence of sibling criminality (SIBCRIM) and negatively associated with age at release (AGEOUT).

The correlations among the regional crime and crime clearance rate variables in Table II.4 have the expected signs. The correlations between both violent and property crime rates (VCRATE and PCRATE) and between violent and property clearance rates (VCLRATE and PCLRATE) are quite strong and positive (r = 0.61 and 0.64 for the crime rates and clearance rates, respectively). Finally, there are weaker (although statistically significant) negative relationships between VCRATE and VCLRATE and PCRATE and PCLRATE.

2.4. Characteristics of Early Failures and Long-term Survivors

As previously noted, the mean time to failure for those who failed during the followup period was 306 days; but, the median of 204 days suggests that the time to failure distribution is highly skewed. (As before, failure is defined as the first arrest/revocation following release.) In this section, the characteristics of two distinct sub-samples of the cohort are compared. Specifically, the characteristics of those who failed very early (within 12 weeks of release) are compared with those who had relatively long-term survival (2 years).

Table II.5 gives the means and standard deviations of the dependent variables for these two groups. For ease of comparison, the values for the complete sample (from Table II.1) are also included. For the most part, the differences in means between the early failures and long-term survivors are consistent with theory and prior research. On average, the early failures were arrested for the first time at a younger age than the long-term survivors. They tend to have significantly more extensive criminal histories in terms of time since first arrest, number of arrests, parole violations, prior commitments, and criminal history scores for prior arrest charges other than violence and robbery. They have worse records of threats, violent behavior and rule breaking during their current commitment. They are also more likely to have had serious gang involvement and school disciplinary problems. Finally, the early failures are more likely to come from backgrounds of poor parenting and to have siblings with criminal records.

More peculiar, perhaps, is the fact that in this cohort the long-term survivors tend (on average) to have been committed for more serious offenses (MWF) than the early failures and to have served longer sentences.

It should be noted that the difference in the mean age at release between these two groups is small and not even close to statistical significance. Both of these groups tended to be slightly, albeit significantly, older when released than the complete population.

There are no statistically significant differences between the early failures and long-term survivors in the measures of mean crime and criminal justice environments, although the higher violent crime rate in counties from which the early failures come (65.39 versus 64.77) just escapes statistical significance as defined in Table II.5 (p(t) = 0.0504).

TABLE II.5. COMPARISON OF EARLY FAILURES AND LONG TERM SURVIVORS[1]

Variable	All	Sample 12-Week Failures	2-Year Survivors
Criminal History:			
AGEFIRST	14.19/(2.81)	13.70/(2.52)	15.33/(2.83)**
CARLNGTH	4.14/(2.56)	4.82/(2.50)	3.21/(2.52)**
NOARRSTS	7.58/(4.64)	9.21/(5.01)	5.40/(3.86)**
PPARVIOL	1.03/(1.43)	1.29/(1.44)	0.33/(0.88)**
PRCOMMIT	1.17/(1.20)	1.38/(1.28)	0.77/(0.98)**
Scores:			
VIOLENCE	1.22/(1.43)	1.24/(1.43)	1.14/(1.40)
ROBBERY	0.58/(0.87)	0.62/(0.91)	0.57/(0.82)
BURGLARY	1.66/(1.71)	1.86/(1.71)	1.32/(1.70)**
OTHPROP	1.33/(1.56)	1.71/(1.84)	0.91/(1.33)**
GENDELQ	3.24/(2.83)	3.73/(2.93)	2.19/(2.41)**
Current Commitment:			
MWF	2.39/(0.52)	2.39/(0.51)	2.53/(0.51)**
CYAVIOL	0.85/(1.30)	0.99/(1.35)	0.74/(1.22)**
INFRRATE	0.82/(1.18)	1.06/(1.30)	0.60/(0.91)**
TIMEIN	1.13/(0.61)	1.12/(0.57)	1.23/(0.65)*
AGEOUT	19.45/(1.84)	19.64/(1.74)	19.77/(1.75)
Substance Abuse and School Problems:			
ALCOHOL	0.84/(0.81)	0.79/(0.80)	0.88/(0.78)
DRUGS	1.02/(0.80)	1.03/(0.77)	0.93/(0.78)
GANG	0.47/(0.79)	0.64/(0.86)	0.33/(0.68)**
DROPOUT	0.55/(0.50)	0.60/(0.49)	0.56/(0.50)
SCHDISC	0.81/(0.82)	0.84/(0.85)	0.68/(0.76)**
Family Background:			
FAMSIZE	0.48/(0.50)	0.46/(0.50)	0.45/(0.50)
FAMVIOL	0.40/(0.79)	0.34/(0.75)	0.42/(0.82)
PARALCH	0.46/(0.80)	0.40/(0.76)	0.42/(0.77)
PARCRIM	0.32/(0.68)	0.32/(0.67)	0.26/(0.61)
SIBCRIM	0.65/(0.86)	0.76/(0.90)	0.56/(0.82)**
WEAKMOM	1.04/(1.00)	1.10/(0.99)	0.85/(0.96)**
Environment:			
PCLRATE	0.17/(0.03)	0.16/(0.03)	0.17/(0.03)
PCRATE	65.24/(10.66)	65.39/(9.51)	64.77/(10.98)
VCLRATE	0.51/(0.09)	0.51/(0.08)	0.51/(0.09)
VCRATE	14.95/(4.49)	15.20/(4.18)	14.57/(4.69)

[1] Sample sizes are N = 1949 (all), N = 360 (12-week failures), and N = 413 (two-year survivors). Asterisks indicate significance of two-tailed t test for significance of difference in means of 12-week failures and two-year survivors: ** $p(t) < 0.01$; * $p(t) < 0.05$.

Finally, regional differences in short-term failure and long-term survival are shown below.

		Failure Rates	
Region	Description	12-Week Failures	2-Year Survivors
LA	Los Angeles	0.220	0.181
BAYAREA	San Fran. Bay	0.135	0.216
SONOTLA	So. CA, not LA	0.195	0.241
NORCNTRL	North/Central	0.159	0.240

This table indicates, for example, that 22 percent of the LA sample failed within 12 weeks while only 18 percent survived more than 2 years. The null hypothesis that chance alone could account for the observed differences among the four Regions either in the early failure rates or in the rates of long-term survival is quite implausible. The chi square for the early failure data is 14.9. With three degree of freedom the probability under the null hypothesis is 0.002. Similar results for the long term survival are a chi square of 8.3 and an associated probability of 0.04.

2.5 Final Data Note

The descriptive statistics presented in this section are intended to convey some sense of the study population from the point of view of a number of variables of interest to criminological theory. Certainly, there are very few theoretical surprises in any of the simple relationships shown here. The analytical task for the remainder of this paper is to assemble the information contained in the data into an "intelligible" score function that can be demonstrated to have some credibility as an assessment of a subject's rearrest risk.

CHAPTER III
MODEL ESTIMATION AND TESTING

3.1 Introduction

A variety of analytic methods have been used to determine how a set of theoretically relevant, independent variables are related to an individual's probability of recidivism (Farrington and Tarling, 1985; Gottfredson and Tonry, 1987). In this paper the analysis of the relationship between a subject's history and his observed outcome (whether or not he was rearrested in the follow-up period) is based on the notion of the hazard function (Kalbfleisch and Prentice, 1980; Elandt-Johnson and Johnson, 1980; Allison, 1984; London, 1988).

Cogent arguments for the particular suitability of hazard analysis to the study of recidivism are given in the literature (Maltz, 1984; Schmidt and Witte, 1988). The two most important advantages of hazard analysis when compared to a "static" model such as a logit are, first, that it makes more efficient use of the data, and second, that information about time to failure has both theoretical and practical importance.

In a model in which the output is the probability of rearrest within some fixed time period T (3 years after release, perhaps), any subject who did not fail but who for one reason or another was not observed for the full time T must be dropped from the analysis. Furthermore, subjects who failed on day T+1 must logically be treated as successes. Neither of these limitations would generally apply under a hazard analysis.

With regard to the question of time to failure, a "static" model necessarily treats a first day failure as an event completely equivalent to a failure on the last day of the defined observation period. It could, however, be theoretically interesting to investigate whether early and late failures come from sub-populations with sensibly different characteristics. From a policy perspective, a realistic question might be whether some treatment program was successful in substantially delaying recidivism, even if it could not claim much success in effecting an overall, long term reduction.

To these well-known arguments we would simply add that the hazard function approach is unique in the flexibility it allows the analyst to address questions of potential importance like the allocation of parole resources or early estimation of the effects of a policy change--a point we hope to make in Chapter VI.

In this chapter we first review briefly the definition of the hazard function and its relation to survival probabilities calculated over finite time intervals. The next section introduces the form of the hazard function used in this paper and outlines the estimation procedure. The final section examines the model's "predictions," comparing various expected outcomes with those actually observed.

3.2 Basic Relations of Hazard Function Analysis[11]

The hazard function, h(t,Z), is defined as a conditional probability density. Specifically, for an arbitrary but short time interval dt, h(t,Z) dt is the probability of rearrest during the time (t, t+dt) of a subject characterized by the covariate vector **Z** under the assumption that he has not yet been

---

[11]Appendix A contains a mathematically more detailed discussion of the hazard function and the estimation procedure.

arrested by time t. With S(t,Z) defined to be his unconditioned probability of survival to t, it follows that

$$dS(t,Z) = -S(t,Z)\,h(t,Z)\,dt \qquad 3.1$$

and, therefore, that

$$S(t,Z) = e^{-\int_0^t h(x,Z)\,dx} \qquad 3.2$$

For present purposes what is important to note here is that all probabilities relating to survival or failure are determined once the hazard function is specified. For example, a relationship that is used rather extensively in what follows is the conditional probability that a subject will be rearrested at some point in the finite time interval $(t_1, t_2]$, given that he has not yet been rearrested by time t_1 -- perhaps in his third year at risk, supposing he has survived arrest-free for two years. From the definitions above, this can be expressed as

$$\begin{aligned} P_c(t_1,t_2,Z) &= \frac{S(t_1,Z) - S(t_2,Z)}{S(t_1,Z)} \\ &= 1 - e^{-\int_{t_1}^{t_2} h(x,Z)\,dx} \end{aligned} \qquad 3.3$$

The log-likelihood function on the data can be written quite generally in terms of the hazard function as

$$\ln L = \sum_{failures}^{F} \ln h(t_i,Z_i) - \sum_{all}^{N} \int_0^{t_i} h(x,Z_i)\,dx \qquad 3.4$$

Here t_i is, of course, subject i's observed time to failure (rearrest) or censoring (no rearrest by the end of the period of data collection). The model is estimated as usual by maximizing ln L with respect to all parameters contained in the function h(t,Z).

3.3 Estimation of the Hazard Function

In this paper we choose to work with a function defined by the log linear relation:

$$\ln h(x,Z) = Z'(c + bx + a \ln x) \qquad 3.5$$

or

$$\begin{aligned} h(x,Z) &= e^{Z'c}\, x^{Z'a}\, e^{Z'bx} \\ &= \prod_1^{K} \left(e^{c_k} x^{a_k} e^{b_k x}\right)^{Z_k} \end{aligned} \qquad 3.6$$

Here, **c**, **b**, and **a** are parameter vectors.[12] The covariate vector **Z** is assumed to have K components. Note that this function factors into a product of terms, each of which depends on only one covariate. The unit of time x is years. Since (h dx) is a conditional probability and, therefore, dimensionless, the hazard function has dimensions of 1/years.

This form was motivated by two assumptions about an individual's recidivism risk. It was first assumed that the risk is relatively low for most (not necessarily all) subjects in the first week or so immediately following release. This suggests that in the very early period the risk might increase as a positive power of the time since release. Second, it was assumed that most subjects who have survived arrest-free for a very long time might now be expected to have low probabilities of failing the next day or the next week. Such a conditional probability might be represented mathematically by including a term that decreases exponentially with time. From this latter choice it also follows that most subjects are assumed to have some finite probability of not being rearrested at all.

While these assumptions motivated the choice of the model's form, they need not hold for all subjects. The function allows for the possibility that for some subjects the initial risk will be high ($\mathbf{Z}'\mathbf{a} < 0$) or that for some the risk appears to continue to increase indefinitely ($\mathbf{Z}'\mathbf{b} > 0$).

Some degree of parsimony was achieved by randomly splitting the data set into two approximately equal halves, separately estimating models on the two halves by likelihood maximization, cross-validating and setting to zero parameters that were "inconsistently" related to failure in the two subsets of observations.[13]

Independent variables were added in successive groups to a model whose <u>form</u> was taken from the previous solution. (That is, parameters once set to zero were defined to be zero in subsequent runs. Other parameters were re-estimated.) The initial model contained the 11 criminal history variables (including the charge for the current commitment) and the intercept term. In the second run 3 variables characterizing the current commitment were added: TIMEIN, CYAVIOL, and INFRRATE. This was followed by inclusion of the 5 delinquency variables (ALCOHOL, DRUGS, GANG, DROPOUT, AND SCHDISC). At the fourth stage, the 6 family pathology variables were added (FAMSIZE, FAMVIOL, etc.) and then the 3 regional variables (LA, BAYAREA, SONOTLA).[14] Next the county violent crime and violent crime clearance rates (VCRATE, VCLRATE) were included, followed by these rates for property crimes. The last variable added was AGEOUT, the subject's age at release.[15]

---

[12]An application of this model to the study of pretrial failure is given in Visher and Linster, 1990.

[13]A brief discussion of informal model selection using cross-validation methods is given in Kmenta, 1986.

[14]See footnote 7 of Chapter II for a listing of counties making up each region. North/Central was treated as the reference category.

[15]The variable INCRIME was defined as: INCRIME = (AGEOUT - AGEFIRST - TIMEIN). These variables stand as surrogates for 4 different theoretical constructs. By definition, however, only 3 of them are <u>mathematically</u> independent. Consequently no more than 3 can appear at a given stage of model identification in any one of the coefficient terms **a**, **b**, or **c**. It was arbitrarily decided to enter AGEOUT last. At this point at least one of the other variables had been dropped from each coefficient term.

A model developed in this way resulted in a statistically good fit as measured by a likelihood ratio test comparison with a model in which it is assumed that the covariates Z contain no reliable information about rearrest risk and further that the risk level for all subjects is constant in time -- that is, a "naive" model in which the hazard function itself is simply a constant.[16] However, a closer examination of the fitted model indicated that it was systematically underestimating the risk in the very early period. It was decided, therefore, to model the rearrest risk as if there were two epochs in the post-release period. The point of division was taken somewhat arbitrarily at 36 weeks.[17] The hazard function used in the analyses of this paper was defined as:

$$h(t,Z) = \begin{cases} e^{Z' c_1} t^{Z' a_1} e^{Z' b_1 t} & t \le 0.6904 \\ e^{Z' c_2} (t-0.6712)^{Z' a_2} e^{Z' b_2 (t-0.6712)} & t > 0.6904 \end{cases} \qquad 3.7$$

where t is years after release.[18]

The model coefficients and their estimated t statistics are given in Table III.1[19] Those covariates followed by asterisks (***) were dropped in the course of model identification. The implication is that in this population they have little independent power to discriminate between risk levels. For future reference it might also be noted that the model for the later period is relatively parsimonious in comparison to that for the first 36 weeks. With fewer variables reliably related to subsequent failure, given 36 weeks survival, the model will obviously have less power to discriminate between risk levels of individual subjects.

The exploration of the role the different covariates play in the assignment of risk is the subject of a later chapter. Here we would emphasize only that this model is a mathematical representation that purports to discriminate between rearrest risk levels of subjects only in a population very similar to the one from which these data were drawn. Even a cursory examination of the entries in Table III.1 suggests that a quite different model might have resulted had the population of interest been chosen differently. For example, the number of arrests, the charge score for violence, a history of family violence or of parent criminality might all have quite strong risk-discriminating power in a population defined as delinquent but without the depth of criminal experience that characterizes most of this CYA release cohort. Here none of these variables were found to contribute significant and reliable information about rearrest risk over and above that contained in other, presumably correlated terms remaining in the model.

---

[16] A likelihood ratio test, comparing the fitted model with a "naive" model, resulted in a chi squared value of 786. With 38 degrees of freedom the probability under the null hypothesis is less than 10^{-9}.

[17] About half the 1949 subjects (979) failed within the first 36 weeks. Among the total of 1710 subjects rearrested during the entire follow-up period the mean time to failure was 306 days with a standard deviation of 293 days. The median was considerably smaller, 205 days, indicating a heavy concentration of fairly short failure times and a long tail to the right. The coefficient of skewness of this distribution is 1.55.

[18] 36 weeks = 0.6904 years. The form for the later period was chosen so as to exclude the singularity at 0 from the range of definition. See Appendix A.

[19] See Appendix A for a brief discussion of the different roles of the **a**, **b** and **c** coefficient vectors. Note that the variables **Z** are not standardized. Consequently, the coefficients are not dimensionless and their magnitudes cannot be compared directly.

TABLE III.1 MODEL SOLUTION[1]

Variable	Model Coefficients (Estimated t statistic) t < 0.690 years			t > 0.690 years		
	c_1	b_1	a_1	c_2	b_2	a_2
AGEFIRST		0.145 (2.03)		-0.044 (4.96)		
INCRIME	0.056 (2.17)					
NOARRRSTS	***					
PPARVIOL	0.176 (4.35)		0.038 (1.53)	0.277 (9.06)		0.034 (1.63)
PRCOMMIT			-0.055 (3.33)	0.059 (1.65)		
CH Scores:						
VIOLENCE	***					
ROBBERY		0.429 (3.70)				
BURGLARY		0.219 (3.55)	-0.018 (1.34)	-0.027 (1.18)		
OTHPROP	0.098 (4.87)			0.061 (2.52)		
GENDELQ	0.105 (4.33)		0.032 (2.12)		0.041 (3.04)	
MWF	0.762 (1.66)	-1.69 (2.20)	0.267 (1.63)			
CYAVIOL	0.123 (2.59)		0.042 (1.43)			
INFRRATE			-0.040 (2.46)			
TIMEIN				-0.118 (1.93)		
AGEOUT	-0.356 (3.04)	0.417 (2.08)	-0.169 (2.46)			
DRUGS			0.036 (1.33)			
ALCOHOL	-0.071 (1.59)					
GANG			-0.090 (3.41)			-0.044 (1.26)
DROPOUT	-0.787 (1.45)	1.51 (1.72)	-0.358 (1.80)			
SCHDISC					0.086 (2.06)	
FAMSIZE			0.077 (1.89)	0.211 (2.81)		
FAMVIOL	***					
PARALCH	-0.071 (1.69)					
SIBCRIM			-0.041 (1.74)		-0.099 (2.37)	
PARCRIM	***					
WEAKMOM			-0.035 (1.69)		0.067 (2.04)	

(CONTINUED)

TABLE III.1 (CONTINUED)

Variable	Model Coefficients (Estimated t statistic) t < 0.690 years c_1	b_1	a_1	t > 0.690 years c_2	b_2	a_2
CJ Envir.:						
PCLRATE					2.74 (2.21)	
PCRATE		0.032 (2.55)				
VCLRATE	1.61 (3.19)					
VCRATE	-0.087 (2.41)			0.023 (2.95)		0.007 (1.91)
Region: (NORCNTRL = reference category)						
LA	0.815 (4.09)		0.221 (2.40)		0.210 (2.35)	
BAYAREA		0.765 (2.69)				
SONOTLA ***						
Intercept	5.09 (2.15)	-10.70 (2.73)	3.41 (4.00)		-1.20 (4.75)	
N		1949			970	
Ln L		-792.4			-1035.2	
Comparison with naive model:						
df		38			17	
Null hypothesis probability:		$< 10^{-10}$			1.5×10^{-9}	

1 Maximum-likelihood solution for the two-period hazard model of equation 3.7; see Table II.1 for definitions of variables. *** indicates variables that were not retained in the final solution.

3.4 Goodness of Fit -- Observed vs. Expected Numbers of Rearrests

While the likelihood ratio test results given at the bottom of Table III.1 show that the covariates contain much significant information, they do not necessarily imply that the model is in fact a "good" one. Conceivably it is simply much better than a very bad, "naive" model. In this section some simple tests are applied to assess in a more concrete way how much agreement exists between some of the model's mathematical consequences and the outcomes actually observed.

Figure III.1 shows the cumulative number of observed first rearrests as a function of time, aggregated over a sequence of four-week intervals. For each point the expected number of first rearrests is calculated by summing the probability of failure by the end of that interval over all 1949 subjects. The

2 standard deviation band about the expected failures curve is calculated at each point T_k in the usual way for a mixed binomial:

$$SD(T_k) = \sqrt{\sum_{i=1}^{N}\{P_i(T_k)(1-P_i(T_k))\}} \qquad 3.8$$

where

$$P_i(T_k) = 1 - e^{-\int_0^{T_k} h(x,Z_i)dx} \qquad 3.9$$

The maximum discrepancy between the expected and observed accumulated failures is 22, occurring in the interval ending with week 52. Overall during the first 160 weeks 1664 subjects were rearrested at least once; the expected number projected by the model is 1677.8.

While Figure III.1 gives some idea of the decreasing rate at which failures are accumulating, considerably more information about the model's fit to the observed data is contained in Figure III.2. Here the number of observed first rearrests in each 4 week interval is shown along with the number of failures expected among the population surviving to the beginning of the interval. For each subject, therefore, the intervals represent a sequence of Bernoulli trials, terminating with the interval in which he fails or with interval 40 if he is not rearrested within 160 weeks.

Let $P_c(T_k,Z_i)$ be the probability that subject i will fail during the k^{th} 4 week interval, conditioned on his having survived to the beginning of that interval. The expected number of failures is in each case the sum of P_c over all subjects who have not yet been rearrested and, thus, are still at risk at the beginninng of interval k. The standard deviation of the expected failures is again the square root of the sum of $P_c(1 - P_c)$.

From these definitions it follows that the quantities

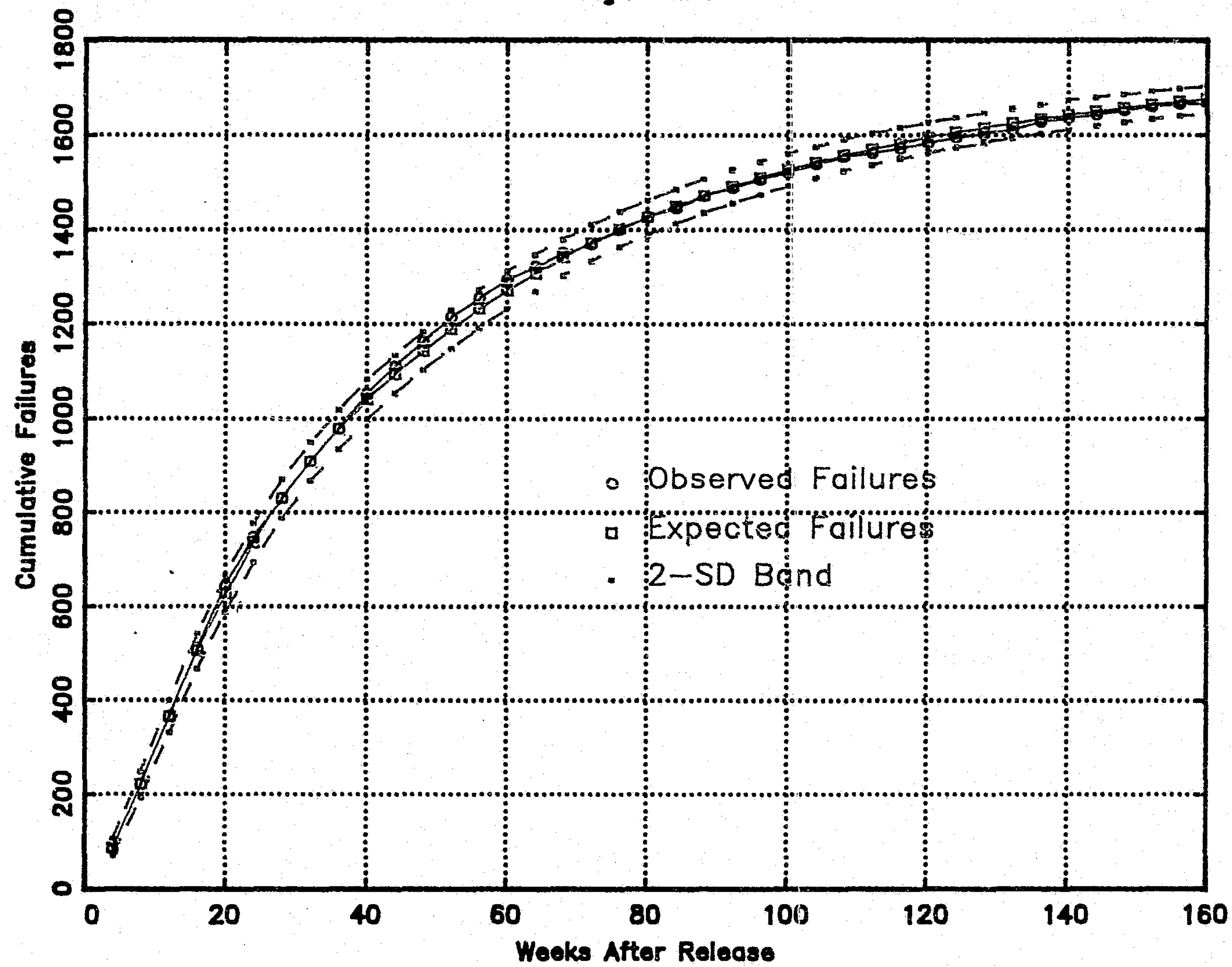
$$z_k = \frac{([Expected\ rearrests]_k - [Observed\ rearrests]_k)}{SD_k} \qquad 3.10$$

are independent for different k and, under the null hypothesis, are asymptotically distributed N(0,1). Consequently, the sum of their squares is asymptotically chi-square distributed with degrees of freedom equal to the number of intervals. Here this sum equals 34.74. With 40 degrees of freedom the probability under the null hypothesis is 0.706.[20] This result may be interpreted somewhat loosely as saying that the average four-week failure probability that the model assigns to survivors is, as a function of time, in reasonably good agreement with what is observed.

We next turn to the question of how well the model discriminates risk levels among individuals. Figure III.3a shows the distribution of the probability of a rearrest within three years for the entire population and for two sub-populations identified with the benefit of hindsight: those 979 subjects who actually failed within the first 36 weeks and the 413 actually surviving arrest-free for at least 2 years. Clearly

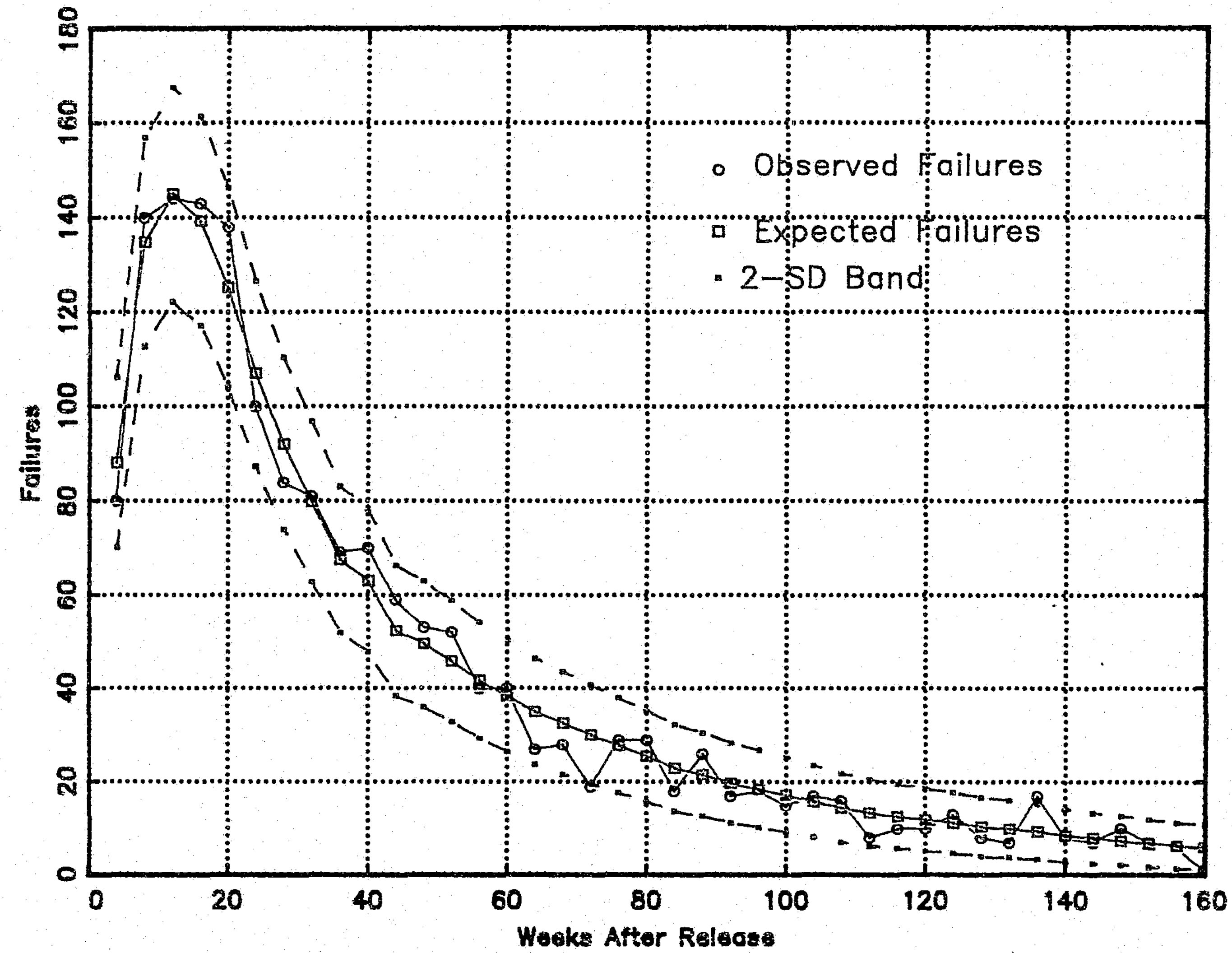
[20]That is, if one choses to accept the model as a valid representation of rearrest as a stochastic process, overall deviations of the order of those shown in Figure III.2 should be expected in about 70% of all applications to similar populations.

Figure III.1



1800
1600
1400
1200
1000
800
600
400
200
0
Cumulative Failures
0
20
40
60
80
100
120
140
160
Weeks After Release
Observed Failures
Expected Failures
2—SD Band

Figure III.2



Observed Failures
Expected Failures
2—SD Band
Failures
0
20
40
60
80
100
120
140
160
180
0
20
40
60
80
100
120
140
160
Weeks After Release

the modeled probabilities are in qualitative agreement with the observed outcomes: the early failures are assessed as posing a significantly higher risk than 2 year survivors. But it should be noted that no group is identified that might be considered to be low risk in an absolute sense. The minimum 3 year failure probability assigned to any subject is 0.48. Even among those subjects who turned out to be 2 year survivors, the initial 3 year failure assessment has a median probability of 0.76.

Figure III.3b examines the results from a somewhat different perspective. What is plotted here is the distribution of probabilities of surviving through the third year at risk, conditioned (hypothetically) on surviving at least the first 2 years. Among the group of early failures, even had they survived for 2 years, the model is not overly optimistic about the chances of third year success. Their median conditional survival probability is only 0.56. The median for the group not rearrested during the first 2 years at risk is 0.73.

A statistically more precise evaluation of how accurately failure probabilites are assigned by the model can be obtained by dividing the probability range into non-overlapping intervals and using these intervals for "classifying" subjects. In Figure III.4a this classification is again based on the probability of failure within three years after release. The intervals are of width 0.05. Thus, the points on the graphs at x = 0.625, for example, correspond to those subjects whose assigned three year failure probability is greater than 0.60 and less than or equal to 0.65.

For each interval the expected numbers of failures and the standard devaitions were calculated in the usual way for a mixed binomial distribution. To simplify the interpretation of the results, expected and observed failures were normalized by dividing by the number of subjects assigned to the interval. Thus, the expected failure "rate" is by definition virtually a straight line through the origin with slope 1.

The distribution of the population over the intervals increases monotonically. Only 1 subject falls into the interval (0.45, 0.50]; only 9 into the interval (0.50, 0.55]. Almost 28% of the 1949 subjects are classified as having a 3 year failure probability greater than 0.95.

When forecast over three years, the agreement between the distributions of expected and observed failures is clearly not very good. If the 10 subjects in the two lowest probability intervals are combined, there are 10 degrees of freedom and the value of chi-square is 20.9. The associated probability under the null hypothesis is 0.02. Of even more significance, however, are the over-estimation of failure probabilities among lower risk subjects and the somewhat less obvious under-estimation among higher risks. These are apparently a reflection of systematic errors in the assignment of 3 year failure probabilities.

Instead of trying to predict failure within three years, suppose we restrict attention to the early period after release, analyzing the data in the same way but comparing expected and observed failure rates during the first 36 weeks. The results are shown in Figure III.4b. The systematic errors now seem much less severe. Furthermore, combining the lowest two intervals, the value of chi-square is 15.65. With 17 degrees of freedom, the null hypothesis probability is a satisfactory 0.55.

If the population under consideration is the 970 subjects surviving 36 weeks at risk and the "classification" is based on the conditional probability of failure within the first 3 years, given 36 weeks survival, the systematic errors return; and with 13 degrees of freedom, chi-square has the unsatisfactorily large value of 32.8. However, if the population of interest is limited still further to the 413 subjects surviving 2 years, with the "classification" based on their conditional probability of failure during the 3rd year, the systematic errors again seem to disappear; chi-squared equals 9.2; there are 14 degrees of freedom; and the asymptotic probability under the null hypothesis is 0.82.

Suppose there exists within the population a set of offenders that might loosely and somewhat fancifully be described as "chronic" revidivists -- subjects who can be relied on to fail and to do so relatively early (here, 36 weeks) in the period after release. The 36 week results might be interpreted as indicating that the model's covariates appear to be doing a reasonable job of assigning to all subjects a probability of membership in this "chronic" group.

For the remainder of the population (the 36 week survivors), a prediction of failure is more problematic. Certainly, the systematic discrepancies between observed and expected failure rates noted above are consistent with a hypothesis that some 36 week survivors have characteristics that indicate a risk level lower than that determined only by their arrest histories and other covariates considered in the model for the later period. Particularly at the lower end of the risk scale there may be a heterogeneity among subjects that could be "explained" by variables not included here. The discriminating power of such variables does not seem to last very long, perhaps because their variance becomes small as the population of survivors shrinks in time.

This is speculative, of course; but it is a point to which we shall return in a later chapter.

Figure III.3a

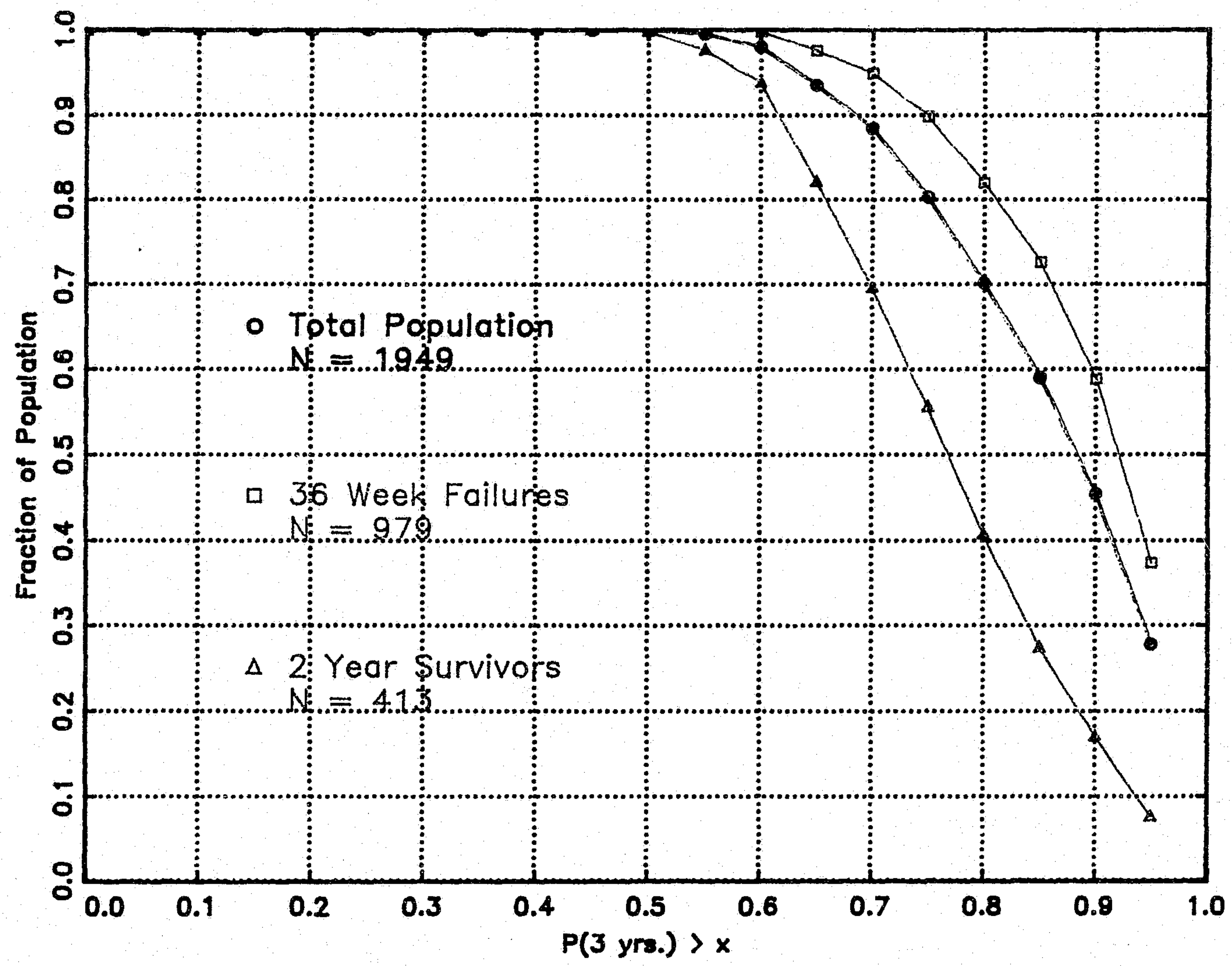


Figure III.3b

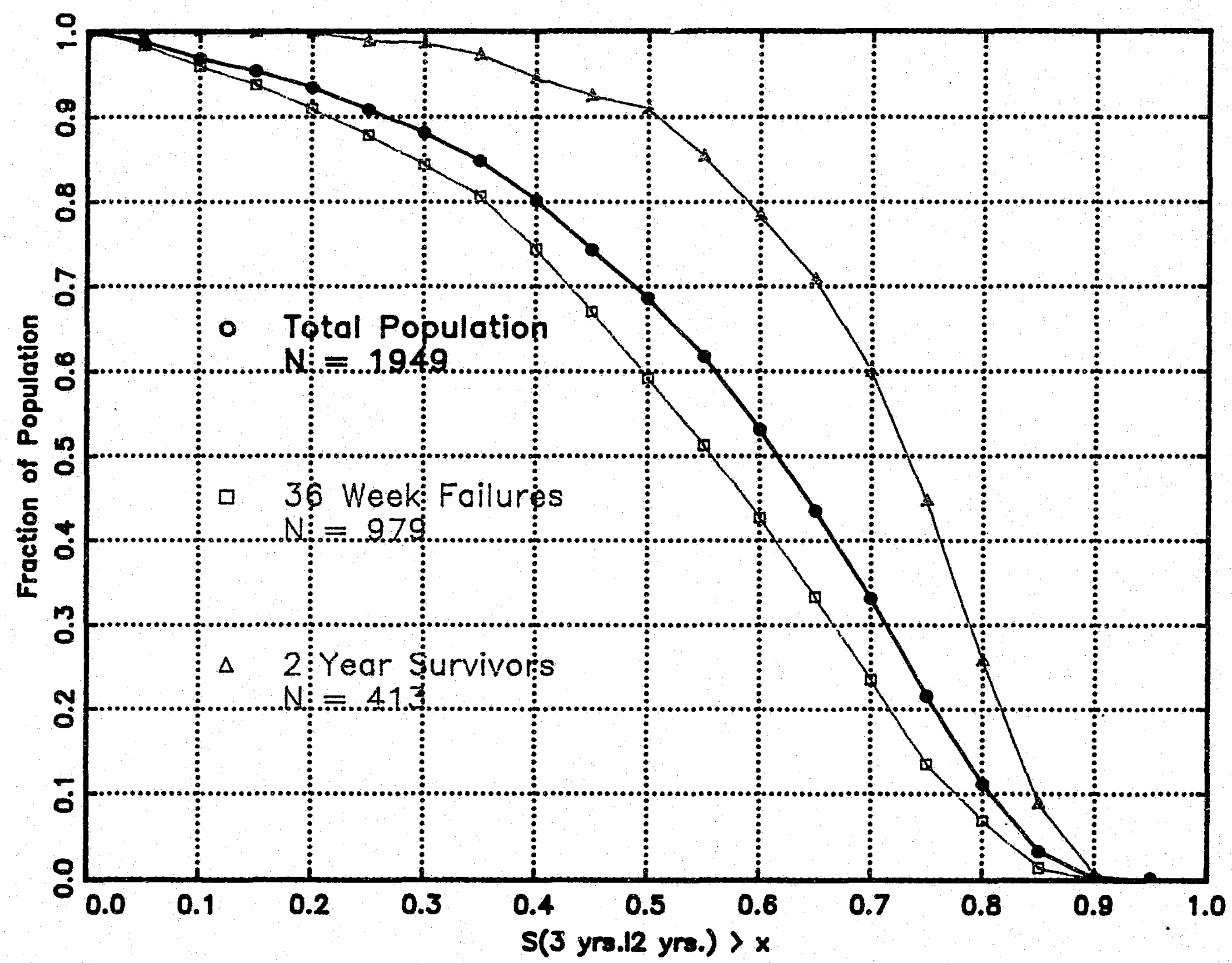


Figure III.4a

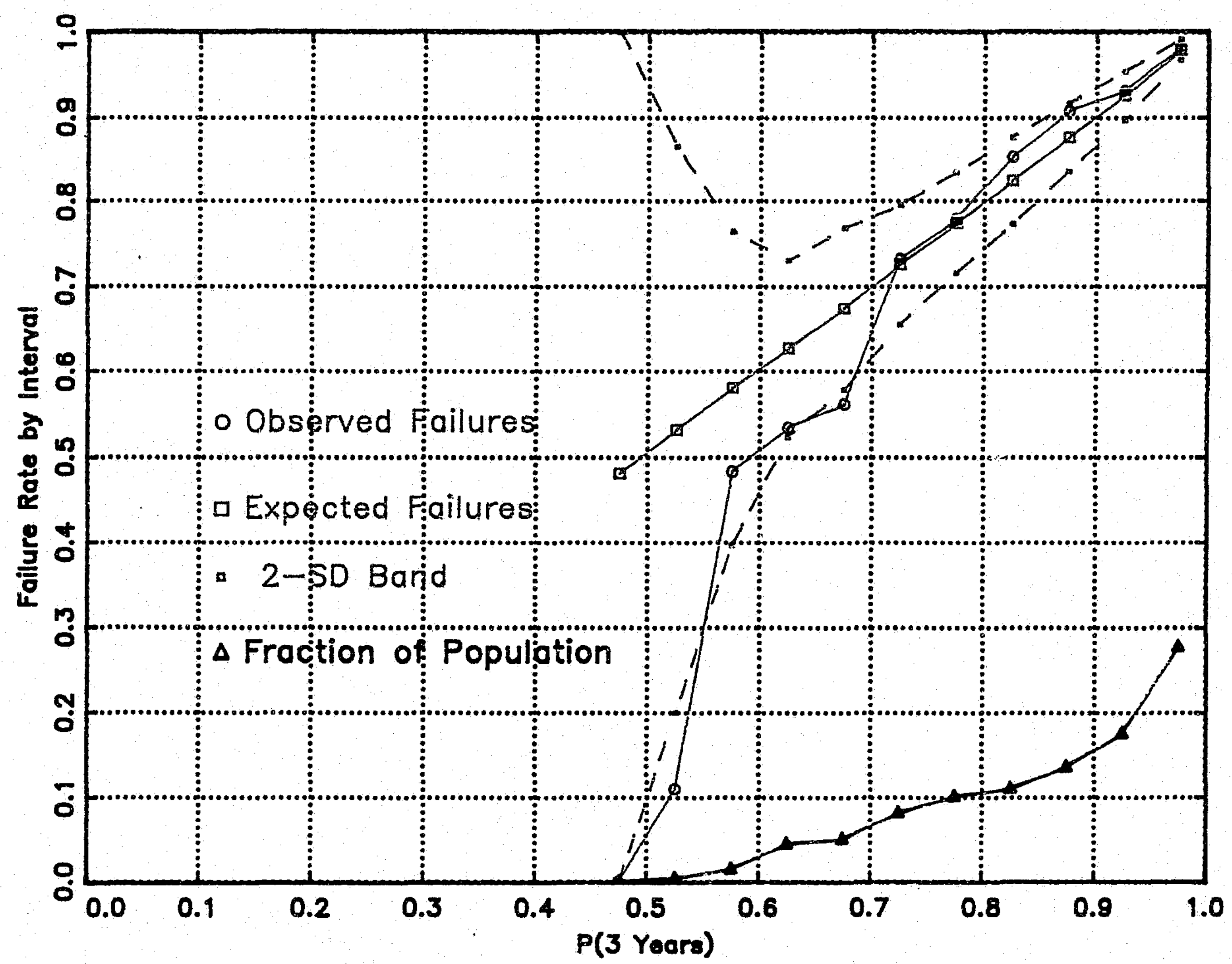
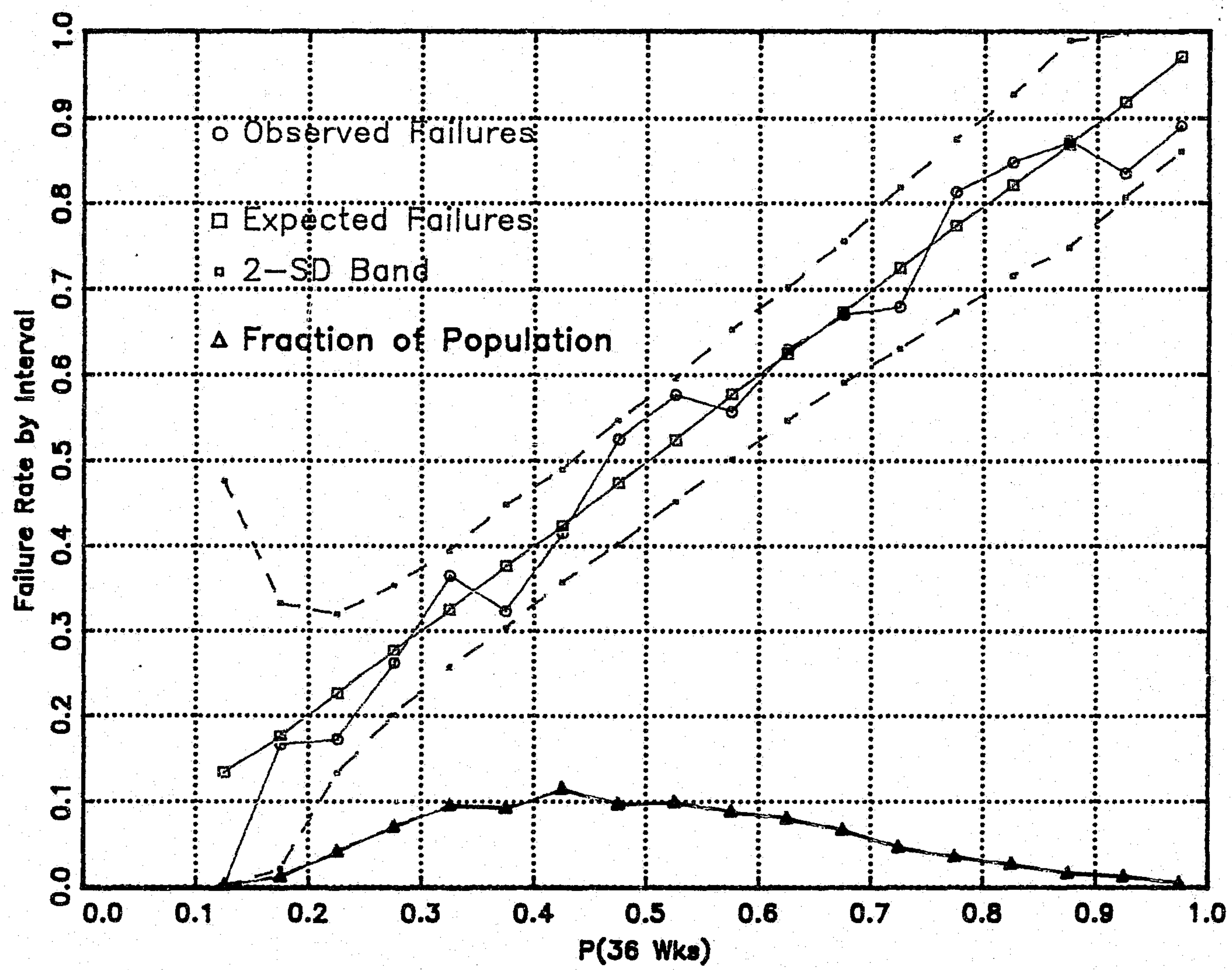


Figure III.4b



CHAPTER IV
MODEL INTERPRETATION

4.1. Introduction

The previous chapter presented the hazard model and its maximum-likelihood solution. One can consider a probability model simply as the "black box" solution of an empirical, likelihood-maximization problem. But it is tempting to inquire into its inner workings, analyzing its handling of the complex of inter-variable correlations and seeking some measure of the relative strength of the independent variables in the model's determination of subject risk levels.

In this chapter, we continue to explore the implications of the model in terms of the impact of covariates on failure rates. First, we posit a comparison of hypothetical pairs of subjects who differ only with respect to the value of one variable--a ceteris paribus analysis. Subsequently, we reject this analysis as unrealistically simplistic, acknowledging that subjects who differ on one variable are likely to differ to a predictable extent on others. We take advantage of the descriptive statistics presented in Tables II.1, 3 and 4 to examine the risks ascribable to hypothetical samples of subjects.

Each line of the model of Table III.1 in effect defines a covariate's contribution to rearrest risk as a function of time since release. The covariates are not standardized variables; consequently the coefficients are not dimensionless and their magnitudes within any column cannot be compared directly. Qualitatively, one can say that in the very early period after release covariates with negative valued coefficients $\mathbf{a}_1$ tend to be associated with high initial risk; for long term survivors higher residual risk is associated with covariates having positive valued coefficients $\mathbf{b}_2$.

To put things in more quantitative terms, we consider the relative risk posed by a pair of subjects, i and j. Let

$$\Delta \mathbf{Z} = \mathbf{Z}_i - \mathbf{Z}_j$$

be the vector of covariate differences. Since the hazard functions used in this paper are log linear in Z, their ratio factors into a product of terms, each depending on only one component of $\Delta \mathbf{Z}$:

$$\frac{h_i(t,Z_i)}{h_j(t,Z_j)} = \prod_1^K (e^{c_k} t^{a_k} e^{b_k t})^{\Delta Z_k} \qquad 4.1$$

where K is the number of covariates in the model.

The simplest comparison of relative hazards would examine the ratio of risk of two subjects i and j whose covariate vectors are identical except for a difference of 1 on a single component--say covariate k.[21]. In this case, equation 4.1 reduces to:

---

[21] Such a difference is, of course, impossible by definition for the two clearance rate variables, where a difference of 0.1 is in fact quite extreme.

$$\frac{h_i(t,Z_i)}{h_j(t,Z_j)} = e^{c_k} t^{a_k} e^{b_k t} \qquad 4.2$$

Values greater than 1 imply that subject i poses a greater relative risk at time t than subject j.

Table IV.1 shows the results of evaluting equation 4.2 for each of the K covariates at 2, 52, and 104 weeks.[22] Each line of Table IV.1 gives some idea of how the model's assessment of a variable's discriminating power is changing in time. For example, consider the entries for AGEFIRST. During a short initial period after release the subject who was older at the time of his first arrest is considered to be a higher risk than his otherwise identical counterpart. But among survivors of the initial period a higher risk is assigned to the subject whose arrest career started earlier. Or, to take another example, the subject with a higher number of prior parole violations (PPARVIOL) is throughout the follow-up period considered the higher rearrest risk. The discriminating power of this variable increases substantially as time goes on and the population still at risk grows smaller.

Table IV.1, however, is not very helpful in assessing the relative power of different variables. The problem is that it provides no empirical basis for deciding how much of a difference in the k^{th} variable should be assigned for the comparison of the two hypothetical subjects who are "otherwise identical". A second problem with the "otherwise equal" approach is that as the analysis is extended to greater differences in a variable's value for a pair of subjects, it becomes less plausible that the subjects would be identical on other measures. Thus, for example, an individual who was much younger at the time of his first arrest (i.e., a low value for AGEFIRST) would be expected to have been engaged in crime for a longer period of time (i.e., a higher value for INCRIME) since the sample correlation between these two variables is -0.77 (see Table III.3).

The next section describes a different analytic approach to the comparison of the model's variables. Following this, Section 3 presents results showing the "net effect on risk" for each variable and analyzing this "net effect" for a few selected variables. Finally, Section 4 defines a summary statistic that characterizes in a broad fashion the role that each variable plays in the model's assignment of risk.

---

[22]If we were interested in some ΔZ_k not equal to 1, we could obtain the relative risk by raising the entries in Table IV.1 to the power of $\Delta Z_k = (Z_{ki} - Z_{kj})$.

TABLE IV.1. VARIABLE-SPECIFIC RELATIVE HAZARD CONTRIBUTIONS[1]

Variable	Weeks After Release: 2	52	104
AGEFIRST	1.01	0.96	0.96
INCRIME	1.06		
NOARRSTS			
PPARVIOL	1.05	1.27	1.33
PRCOMMIT	1.20	1.06	1.06
CH Scores:			
VIOLENCE			
ROBBERY	1.02		
BURGLARY	1.07	0.97	0.97
OTHPROP	1.10	1.06	1.06
GENDELQ	1.00	1.01	1.06
MWF	0.84		
CYAVIOL	0.99		
INFRRATE	1.14		
TIMEIN		0.89	0.89
AGEOUT	1.24		
DRUGS	0.89		
ALCOHOL	0.93		
GANG	1.34	1.05	0.99
DROPOUT	1.55		
SCHDISC		1.03	1.12
FAMSIZE	0.78	1.23	1.23
FAMVIOL			
PARALCH	0.93		
SIBCRIM	1.14	0.97	0.88
PARCRIM			
WEAKMOM	1.12	1.02	1.09
CJ Envir.:			
PCIRATE		2.44	37.56
PCRATE	1.00		
VCLRATE	5.00		
VCRATE	1.00	1.02	1.03
Region: (NORCNTRL = reference category)			
LA	1.10	1.07	1.32
BAYAREA	1.03		
SONOTLA			

1 Entries are $(e^{c_k} t^{a_k} e^{b_k t})$

where c_k, a_k, and b_k are given in Table III.1. These entries correspond to the relative hazard risk posed by two hypothetical subjects who differ by the value of 1 on only the k^{th} variable; they are identical on all other variables. Entries left blank correspond to variables not appearing in the model for that time period. Variable definitions are given in Table II.1.

4.2 Explanation of the Analytic Approach

The hazard ratio method (equation 4.1) of analyzing particular variables' effects on subject risk is straightforward and consistent with the mathematical form of the model. But it does not offer much insight into "why" some variables are more important than others in the assignment of risk. While it may be conceptually possible for two subjects to differ on only one variable (except in the case of the chronological variables: age at first arrest, career length (INCRIME), time served and age at release), such pairs may not be like anything actually observed in the data. For example, it would not be implausible to consider otherwise identical subjects who differ by 1 in their number of prior commitments or parole violations and so obtain a sense of the direction and magnitude of change in risk that the model ascribes to such differences. But clearly this cannot be extended very far without creating unlikely hypothetical types. Actual pairs of subjects differing by 2, say, in their number of prior commitments will almost inevitably be observed to have accompanying differences in other criminal history variables.

In what follows, we propose a different approach. Suppose for any given time t we sum the log hazard function over the population and divide by the number of subjects. Because ln h is linear in the covariates, we can write the result as

$$\ln h(t,<Z>) = <Z>'(c + bt + a\ln t) \qquad 4.3$$

Vector multiplication is understood. Here <Z> designates a vector of arithmetic means of the covariates appearing in the model. The function h(t,<Z>) defined by this equation is easily seen to be the population geometric mean of the hazard function at time t.

We are interested in how the geometric mean hazard function varies with changes in the covariate means. Approximating by differentials, we obtain

$$\frac{dh(t,<Z>)}{h(t,<Z>)} = d<Z>'(c + bt + a\ln t) \qquad 4.4$$

The left hand side here is the fractional change in the geometric mean hazard function at time t corresponding to a vector of changes d<Z> in the arithmetic means of the covariates.[23] Thus, instead of comparing two hypothetical subjects, as we did in the previous section, we have here a hypothetical population to be compared to the observed.

In this hypothetical population, we assume that an increase $d<z_k>$ in the mean of the k^{th} variable would realistically be accompanied by changes $d<z_j>$ in the means of other covariates.[24] We define the components of d<Z> in terms of an arbitrary $d<z_k>$ by

---

[23]The validity of the differential approximation requires that the inner product on the right hand side be "small." This does not necessarily mean that all components of <dZ> are "small."

[24]Note that here subscripts denote particular variables of the model and not, as in the previous section, particular subjects.

$$d<z_j> = \frac{s_j}{s_k} r_{jk} d<z_k> \qquad 4.5$$

where s_j and s_k are the standard deviations of z_j and z_k observed in the data and r_{jk} is their correlation. It will be recognized that what we are postulating here is a hypothetical population in which the variable means take on the values they would have if we regressed each z_j against z_k. For each j and k, therefore, $d<z_j>$ is defined to be the change in $<z_j>$ that is "explained" by a given $d<z_k>$, assuming that z_j and z_k are linearly related. For example, in a population in which the mean number of arrests is greater by $d<z_k>$ we would also expect to find that the mean career length is greater by an amount

$$d<z_j> = \frac{2.56}{4.64} \times 0.66\, d<z_k> = 0.36\, d<z_k> \text{ years} \qquad 4.6$$

The following heuristic argument provides a justification for this approach as a way of investigating the relative importance of the different independent variables. Suppose the observed population is a faithful copy of a very large parent population in terms of variable means, standard deviations and correlations. Now imagine drawing from the parent population a long sequence of random samples, each of relatively small size n, and postulate that the regression relations determining the $d<z_j>$'s would hold in the sense of expected values in this sequence. The fractional change in the geometric mean hazard function is then also to be understood as the expected value of increased risk associated with those samples in which the mean of z_k happens to differ from the population mean by $d<z_k>$.

The constraint imposed by the relation between the four chronological variables requires that at least one of the four be handled differently so that the sum of changes in age at first arrest, time in crime, and length of current commitment will always add up to the change in age at release. In what follows, for all $d<z_k>$ except d<INCRIME>, we have arbitrarily chosen to satisfy the constraint by defining

d<INCRIME> = d<AGEOUT> - d<AGEFIRST> - d<TIMEIN>.

When k refers to time in crime, we, again arbitrarily, define

d<AGEOUT> = d<AGEFIRST> + d<INCRIME> + d<TIMEIN>.

Equations 4.4 and 4.5, thus, provide a framework within which to compare the risks posed by hypothetical populations with that posed by our observed population. The hypothetical populations are generated by assuming that the linear relationships between covariates are consistent with those in the observed population as defined in equation 4.5. What remains is to define $d<z_k>$--the change in the k^{th} variable that induces the changes in the $d<z_j>$. We choose to set $d<z_k>$ proportional to s_k so that the different samples are comparable in the sense that these hypothesized differences of sample means are equally probable for all k. Specifically, we assume that the k^{th} variable mean differs from that of the observed population by the value

$$d<z_k> = \frac{s_k}{2} \qquad 4.7$$

The values of $d\langle z_j \rangle$, $j \neq k$, are calculated for the remaining K - 1 covariates using equation 4.5. The fractional change in the geometric mean hazard function at time t is then calculated using equation 4.4.

The result of equation 4.4 is the net fractional change in the geometric mean hazard function that results from increasing the mean of the k^{th} variable, $\langle z_k \rangle$, by $d\langle z_k \rangle$. The net fractional change is the sum of the direct change in the hazard produced by $d\langle z_k \rangle$ and the set of induced changes brought about by the correlated changes in the means of the model's other independent variables.[25]

Table IV.2 presents the results of sequentially allowing each variable to assume the role of the "population-defining" k^{th} variable. Note that the entries in Table IV.2 are in terms of percentage change rather than fractional change (i.e., are the values obtained by equation 4.4 multiplied by 100).

4.3 Interpretation of the Results

Each line of Table IV.2 may be thought of as a measure of the relative power of the model to discriminate between "high" and "low" risk groups if the only information available were individual measures on the single variable z_k -- along with a knowledge of all variables' standard deviations and correlations. For example, if age at first arrest (AGEFIRST) were the only subject-level piece of information we had, the model would consistently ascribe a lower mean risk to a sample rich in late starters. The ability of this one variable to differentiate between risk levels is initially rather modest (-5.6% at 2 weeks) but increases considerably when the focus of attention is recidivism risk among long term survivors (-19.3% at 3 years).

The relative magnitudes of the entries in Table IV.2 provide evidence that throughout the observation period the model assigns risk levels primarily on the basis of a subject's criminal justice record. Most of the other variables may be thought of as adding some discriminating fine structure, distinguishing between subjects with similar criminal histories.

The most striking exception to this is the considerable increase in risk in the period immediately after release that is attributed to a sample that is either older on average or contains a larger fraction of school dropouts[26]. At the time of release the mean age of all subjects was 19.5 years; 54.6% of the population were school dropouts. In the first 4 weeks after release, the average age of the 80 subjects who were rearrested was 20.2 years and 71.3% were recorded as having dropped out of school. This is in some contrast to the group of 138 subjects arrested between weeks 16 and 20, for example. Here the average age was 19.3 and 48.6% were dropouts.

---

[25] It is understood that if the kth variable does not enter the model directly (i.e., all coefficient values are zero, see Table III.1) that the direct change is 0 and that the net change will be equal to the sum of the induced changes.

[26] The correlation between age at release and school dropout is 0.303. In this generally youthful population these variables are measuring similar but not identical constructs.

TABLE IV.2. CHANGE IN GEOMETRIC MEAN HAZARD GIVEN INCREASE $d<z_k>$ IN MEAN OF z_k [1]

Variable z_k	$d<z_k>$	Net % Change in Geometric Mean Hazard at t = Weeks Since Release: 2	16[2]	78	156
AGEFIRST	1.4	-5.6	-13.1	-14.5	-19.3
CARLNGTH	1.3	17.8	12.4	(13.2)	(16.3)
NOARRSTS	2.3	(16.3)	(19.0)	(15.5)	(22.1)
PPARVIOL	0.7	8.3	13.8	23.1	28.1
PRCOMMIT	0.3	14.5	11.2	9.8	13.0
Criminal History Scores:					
VIOLENCE	0.7	(1.9)	(2.1)	(2.3)	(3.0)
ROBBERY	0.4	-1.8	3.3	(-0.4)	(-1.5)
BURGLARY	0.9	7.1	10.1	2.1	3.4
OTHPROP	0.8	10.5	11.8	8.7	10.4
GENDELQ	1.4	6.4	14.0	13.0	22.0
MWF	0.3	-5.5	-5.9	(-4.9)	(-7.5)
CYAVIOL	0.7	3.4	5.8	(0.4)	(1.3)
INFRRATE	0.6	9.9	7.0	(2.1)	(3.4)
TIMEIN	0.3	(-1.0)	(-3.2)	-7.3	-9.0
AGEOUT	0.9	16.0	-3.8	(-6.3)	(-9.7)
ALCOHOL	0.4	0.3	-3.7	(-0.9)	(0.0)
DRUGS	0.4	2.2	1.0	(2.5)	(4.3)
GANG	0.4	9.3	8.0	4.6	5.6
DROPOUT	0.2	14.0	0.4	(-2.0)	(-2.8)
SCHLDISC	0.4	(0.3)	(4.5)	7.4	14.9
FAMSIZE	0.3	-2.8	-1.5	4.3	3.1
FAMVIOL	0.4	(-1.9)	(-1.4)	(0.5)	(1.9)
PARALCH	0.4	-2.4	-1.5	(2.8)	(4.9)
PARCRIM	0.3	(0.8)	(2.4)	(3.0)	(5.2)
SIBCRIM	0.4	7.5	5.8	1.2	-3.0
WEAKMOM	0.5	4.5	4.7	6.4	13.0
PCLRATE	0.02	(-3.2)	(-4.9)	1.3	4.3
PCRATE	5.3	-0.8	0.8	(1.7)	(0.9)
VCLRATE	0.04	1.2	-1.1	(-1.0)	(-0.7)
VCRATE	2.2	-0.4	1.5	4.2	5.7
Region:					
LA	0.2	5.5	7.7	3.4	5.2
BAYAREA	0.2	-4.8	-2.7	(-1.6)	(-2.8)
SONOTLA	0.2	(2.6)	(-2.2)	(-2.4)	(-4.1)

[1] Entries are the results of calculating equation 4.4 with each variable assuming the role of z_k; $d<z_k> = s_k/2$. ( ) indicates that the variable does not appear in the model for this time period. The percent change is entirely "induced" by correlations with variables that are explicitly included.

[2] The geometric mean hazard function passes through its maximum at approximately 16 weeks after release.

One conjecture to explain this very short-lived phenomenon is that there is within the population a small group of very arrest-prone offenders who have both dropped out of school and have reached an age at or near the statutory limit of the Youth Authority's jurisdiction. In effect, then, a relatively few high risk subjects might be "maxing out," with little or no post-release supervision.

As was previously noted, the results shown in Table IV.2 are **net** effects, including both the **direct** effect of change in $<z_k>$ and the **induced** effects of changes in all other variables in the model (the $<z_j>$, $j \neq k$). This may be more clearly seen if we rewrite equation 4.4 as follows:

$$\frac{dh(t,<Z>)}{h(t,<Z>)} = d<z_k>(c_k + b_k t + a_k \ln t) + \sum_{j \neq k} d<z_j>(c_j + b_j t + a_j \ln t) \qquad 4.8$$

$$net = direct + induced$$

Note that the direct effect and induced effects may not necessarily have the same sign and that, as we shall see, for some variables the induced effects are both opposite in sign and greater in magnitude than the direct effect, implying that the direct and net effects have opposite signs.

To illustrate these effects, the direct and induced changes in geometric mean hazard resulting from an initial change in variable $<z_k>$ are presented in Tables IV.3a, b and c.[27] These tables decompose--into contributions made by each variable--the net results reported in Table IV.2 for changes in mean Age at First Arrest (k = AGEFIRST), in mean Time Served (k = TIMEIN), and in fraction of the sample from Los Angeles County (k = LA) at t = 2 weeks and t = 1 1/2 years.

The first line in each of the Tables IV.3 presents the mean of z_k, the percent change in $<z_k>$ defined by $d<z_k>$, i.e. $[(d<z_k>/<z_k>) \cdot 100]$, and the **direct** percentage change in $h(t,<Z>)$ attributable to $d<z_k>$ at t equal 2 and 78 weeks. The remainder of each table enumerates the **induced** percentage changes in the geometric mean hazard attributable to each $<z_j>$. Specifically, the population means $<z_j>$ and the correlations between the z_j's and z_k are shown in the first two columns of Tables IV.3. The next column gives the percent change in the means of the z_j's induced by $d<z_k>$, i.e. $[(d<z_j>/<z_j>) \cdot 100]$. The last two columns show the contribution to the net percent increase in the geometric mean of the hazard function produced by the change in the mean of z_j at 2 weeks and 78 weeks after release. The last line in each Table is the sum of the direct and induced effects and is thus the net effect reported in Table IV.2. A brief interpretation of each table is presented below.

4.3.1 Age at First Arrest (AGEFIRST) (Table IV.3a)

As was seen in Table IV.2, the net percentage change in geometric mean hazard associated with a hypothetical sample 1.4 years older than the observed was negative, implying that this "older sample" posed somewhat less of a risk than that posed by the observed sample. Table IV.3a decomposes this net effect into the direct and variable-specific induced effects. The hypothetical sample that is on average 9.9% (1.4 years) older at first arrest also has other attributes that one might expect. The percentage changes in the $<z_j>$ (induced by $d<z_k>$) included in Table IV.3a provide some insight into

[27] The results of these tables are all derived from linear relations. One might, therefore, equally well consider a sample that is on average 1.4 years younger at first arrest simply by reversing all signs in the percent change columns of Table IV.3a.

TABLE IV.3a. DECOMPOSITION OF THE NET EFFECT OF d<AGEFIRST> ON THE GEOMETRIC MEAN HAZARD[1]

Variable	$\langle z_k \rangle$	$\frac{(d\langle z_k \rangle)100}{\langle z_k \rangle}$	% Change $h(t,\langle Z \rangle)$ t = 2 wks	% Change $h(t,\langle Z \rangle)$ t = 78 wks
AGEFIRST	14.19	9.9	0.8	-6.3

Variable	$\langle z_j \rangle$	r_{jk}	$\frac{(d\langle z_j \rangle)100}{\langle z_j \rangle}$	% Change $h(t,\langle Z \rangle)$ t = 2 wks	% Change $h(t,\langle Z \rangle)$ t = 78 wks
CARLNGTH	4.14	-0.77	-23.8	-5.5	
NOARRSTS	7.58	-0.63	-19.1		
PPARVIOL	1.03	-0.21	-14.6	-0.8	-4.0
PRCOMMIT	1.18	-0.30	-15.4	-3.3	-1.1
Criminal History Scores:					
VIOLENCE	1.22	-0.15	-9.0		
ROBBERY	0.58	-0.03	-2.2	-0.0	
BURGLARY	1.66	-0.28	-14.5	-1.7	0.6
OTHPROP	1.33	-0.24	-13.8	-1.8	-1.1
GENDELQ	3.24	-0.39	-17.1	-0.0	-1.9
MWF	2.39	0.16	1.7	-0.7	
CYAVIOL	0.86	-0.07	-5.3	0.1	
INFRRATE	0.82	-0.09	-6.4	-0.7	
TIMEIN	1.13	0.03	0.7		-0.1
AGEOUT	19.45	0.47	2.2	9.1	
ALCOHOL	0.84	0.09	4.4	-0.3	
DRUGS	1.02	0.02	0.6	-0.1	
GANG	0.47	-0.08	-6.9	-1.0	-0.0
DROPOUT	0.55	0.17	7.8	1.9	
SCHDISC	0.81	-0.22	-10.9		-0.6
FAMSIZE	0.48	-0.03	-1.5	0.2	-0.1
FAMVIOL	0.40	-0.04	-3.7		
PARALCH	0.46	-0.07	-6.3	0.2	
PARCRIM	0.32	-0.14	-15.0		
SIBCRIM	0.65	-0.21	-13.6	-1.2	0.7
WEAKMOM	1.04	-0.19	-9.2	-1.1	-0.5
PCLRATE	0.17	-0.02	-0.1		-0.1
PCRATE	65.24	-0.01	-0.1	-0.0	
VCLRATE	0.51	0.04	0.4	0.3	
VCRATE	14.95	-0.00	-0.0	0.0	-0.0
Region:					
LA	0.40	-0.02	3.9	-0.0	-0.1
BAYAREA	0.20	-0.02	-1.9	-0.0	
SONOTLA	0.19	0.06	-2.9		
Net Effect				**-5.6**	**-14.5**

[1] Blanks correspond to variables that do not appear explicitly in the model for that period (see Table III.1).

how this hypothetical population differs from our observed. They have substantially shorter criminal histories with fewer charges -- especially for property crimes. They are as a group somewhat more likely to have alcohol problems (4.4%) but less likely to have been seriously involved with gangs (-6.9%). They are on average very slightly older when released (2.2%), are more likely to have quit school (7.8%) and generally come from families with fewer problems -- particularly in terms of parents' or siblings' criminality.

In the period immediately after release the "direct" effect of an increased mean age at first arrest on the "riskiness" of the sample is virtually zero (0.8%). The net decrease (-5.6%) is due principally to the fact that such a sample has on average accumulated a less serious criminal record but tends to be somewhat older at release. The contributions these variables make to the net reduction in mean risk are -12.3% and +9.1%, respectively.[28] Considerably smaller contributions are made by the variables characterizing the current commitment (-1.3%) and variables reflecting the sample's generally less troubled family background (-1.9%).

A year and a half later the survivors of this sample have, according to this model, a net mean risk level 14.5% lower than the survivors from the total population. This is due almost entirely to mean differences in the criminal record variables (-13.7%).

4.3.2 <u>TIMEIN</u> (Table IV.3b)

Time served is of interest because it is to some extent manipulable by policy and may have recidivism implications in terms of a deterrent effect. In the population studied here the mean and standard deviation of the length of the current commitment were 412 days and 222 days (or 1.13 and 0.61 years), respectively. In a hypothetical sample in which the mean time served is 523 days the model's assessment of risk immediately after release is virtually no different (-1.0%) from that of the total population. Indeed, TIMEIN does not appear explicitly in the model for the early period (to 36 months post release), having been dropped as not reliably adding to the risk discriminating power of other covariates.

For survivors of this early period, variability in the length of time served in the current commitment does appear to contain information relating to rearrest risk. And in fact, throughout the post-release observation period the model associates a net effect of reduced risk with a longer sentence -- a result that is obviously consistent with deterrence theory.

As shown in Table IV.3b in the period immediately after release this net effect reflects a 4% mean reduction in risk due to the <u>less</u> extensive mean criminal histories of such a sample, which is just about balanced by a 3% increase in risk due to the fact that on average these subjects tend to be slightly older when released. From the third column of Table IV.3b, it is evident that this hypothetical sample is richer in subjects either with a higher than average record of violence and robbery charges or with worse records of rule breaking and violence during their current commitment. Taking into account the 3.9% increase in the score for seriousness of the commitment offense (MWF) from a population mean of 2.39 to a sample mean of 2.48, one might also conjecture that the hypothetical group considered here contains more subjects <u>currently</u> sentenced for violence or robbery.

---

[28]The -12.3% contribution to risk was calculated by summing the contribution to the percentage change in hazard of the variables AGEFIRST, INCRIME, NOARRST, PPARVIOL, PRCOMMIT, VIOLENCE, ROBBERY, BURGLARY, OTHPROP, and GENDELQ.

TABLE IV.3b. DECOMPOSITION OF THE NET EFFECT OF d<TIMEIN> ON THE GEOMETRIC MEAN HAZARD[1]

Variable	$<z_k>$		$\frac{(d<z_k>)100}{<z_k>}$	% Change $h(t,<Z>)$ t = 2 wks	% Change $h(t,<Z>)$ t = 78 wks
TIMEIN	1.13		26.8		-3.6
Variable	$<z_j>$	r_{jk}	$\frac{(d<z_j>)100}{<z_j>}$	**% Change $h(t,<Z>)$ t = 2 wks**	**% Change $h(t,<Z>)$ t = 78 wks**
AGEFIRST	14.19	0.03	0.3	0.0	-0.2
CARLNGTH	4.14	-0.16	-4.9	-1.1	
NOARRSTS	7.58	-0.14	-4.2		
PPARVIOL	1.03	-0.16	-11.2	-0.6	-3.1
PRCOMMIT	1.18	-0.12	-6.2	-1.3	-0.4
Criminal History Scores:					
VIOLENCE	1.22	0.19	11.2		
ROBBERY	0.58	0.19	14.4	0.1	
BURGLARY	1.66	-0.12	-6.1	-0.7	0.3
OTHPROP	1.33	-0.06	-3.5	-0.5	-0.3
GENDELQ	3.24	-0.19	-8.1	-0.0	-0.9
MWF	2.39	0.36	3.9	-1.6	
CYAVIOL	0.86	0.29	22.2	-0.3	
INFRRATE	0.82	0.22	15.6	1.7	
AGEOUT	19.45	0.15	0.7	3.0	
ALCOHOL	0.84	-0.01	-0.6	0.0	
DRUGS	1.02	-0.02	-0.8	0.1	
GANG	0.47	0.08	6.9	1.0	0.0
DROPOUT	0.55	-0.01	-0.3	-0.1	
SCHDISC	0.81	0.03	1.4		0.1
FAMSIZE	0.48	0.06	3.0	-0.4	0.3
FAMVIOL	0.40	0.05	5.2		
PARALCH	0.46	-0.00	-0.1	0.0	
PARCRIM	0.32	0.00	0.1		
SIBCRIM	0.65	-0.05	-3.0	-0.3	0.2
WEAKMOM	1.04	0.01	0.7	0.1	0.0
PCLRATE	0.17	-0.13	-1.2		-0.4
PCRATE	65.24	0.00	0.0	0.0	
VCLRATE	0.51	-0.06	-0.5	-0.4	
VCRATE	14.95	0.04	0.6	-0.0	0.2
Region:					
LA	0.40	0.12	7.5	0.3	0.5
BAYAREA	0.20	-0.06	-6.4	-0.0	
SONOTLA	0.19	0.03	-3.5		
Net Effect				**(-1.0)**	**7.3**

[1] Blanks correspond to variables that do not appear explicitly in the model for that period (see Table III.1).

These relations are certainly not inconsistent with common sense notions about the type of individuals one would expect to serve longer sentences. But they also illustrate some of the complexities of "explaining" risk -- especially in terms of single covariate categorizations of subjects.

From Table IV.2 we note that, of samples defined to be especially rich in violence or robbery scores or in violence and rule breaking during the current commitment, only the robbery sample has the same sign for net effect on very early or very late risk as the sample serving a longer sentence; and only the rule breakers appear to pose an early risk much different from the population mean. If the composition of the net effects are analyzed for violence or robbery scores, for rule infraction rate or for violence during the current commitment, samples with over-representations of such subjects all served somewhat longer sentences. But these groups also have typically longer and more serious criminal histories -- in contrast to a sample chosen specifically to reflect greater than average TIMEIN.

A clue to what is going on here is offered by the decomposition of the net effect of the MWF variable (data not shown). An increase in seriousness of the charge of the current commitment is accompanied in a sample by a general decrease in length and seriousness of all criminal career variables except for a small increase (1.1%) in the mean violence score and a substantial increase (31%) in the robbery score. Simply put, within this population longer sentences were on average being given for a current charge of robbery or felony violence, pretty much independent of the individual's prior record.

In the population there were 103 subjects with the shortest possible prior records: their current commitment resulted from their first recorded arrest. Eighty of these were charged with violence or robbery or both, often accompanied by other charges. Of the remaining 23, none were charged only with the less serious offenses categorized here as "general delinquency." The average length of sentence was 482 days.

As a group these 103 subjects look quite different from the remainder of the population. They were, of course, considerably older at the time of first arrest--18.2 years vs. 14.0 years for subjects with more than one prior arrest--but roughly the same age at the beginning of their current sentence. They have lower mean scores on delinquency variables, especially gang involvement (0.19 vs. 0.49). Their averages on all but one of the family pathology variables are also lower. The single exception, interestingly enough, is family violence (0.51 vs. 0.39).

During the more than 3 years of post-release observation, 52 of these "novices" were rearrested at least once. This 50% recidivism rate is small only in comparison to the 90% rate of the remaining population. For those who were rearrested, the distribution of most serious charge types at first post-release arrest was roughly the same for the two groups. Under a chi-squared test the differences were nowhere near statistical significance. But the more criminally experienced subjects who were rearrested failed sooner on average (301 days) than did the relative novices (481 days). These results would not seem to offer any reliable evidence one way or the other regarding the hypothesis that longer sentences result in an enhanced deterrent effect since a variety of other factors (e.g., number of prior arrests) appear to be playing a role.

One other statistic might be calculated, relating to a possible deterrent effect of a CYA commitment itself on subjects with at least 2 prior arrests. On average these 1846 subjects were rearrested 7.2 times in the course of a career length of 4.45 years. The mean time between arrests prior to the current commitment was 288 days[29]. This is certainly an overestimate in terms of mean number of days actually at risk between successive arrests. This group had on average 1.2 prior commitments, presumably resulting in periods of local confinement or some form of incapacitative restraint. Thus, the length of time subjects were at risk of being rearrested is less than the career length by an unknown amount and the mean time free on the streets between successive arrests before the present commitment should be less than 288 days.

As already noted, for the 1658 individuals from this group who were rearrested during the post-release observation period the mean failure time was 301 days. This slightly longer arrest-free interval following the end of their current commitment is not inconsistent with an hypothesis of a temporary deterrence effect. But alternative (or more likely, complementary) explanations might lie in post-release supervision effectiveness or even in simple hypotheses about gradual maturation out of some highly arrest-prone behaviors of adolescence.

4.3.3 Los Angeles County (LA) (Table IV.3c)

Table IV.2 indicates that a sample with an over-representation of offenders from Los Angeles County poses a mean rearrest risk that is both slightly higher than that of the total population and fairly constant in time over the course of the follow-up period. The third column of Table IV.3c shows that on average such a population would have criminal records only modestly more serious than those of the total population except for fewer prior commitments, lower general delinquency charge scores and considerably higher scores for robbery charges. Quite possibly what is reflected here is selectivity within the County in juvenile arrest and charging policies and in the use of commitment to the Youth Authority. Both the MWF and TIMEIN variables indicate a marginal increase in mean seriousness of the commitment offense.

While the net increase in mean risk in this hypothetical sample is fairly small and about the same in the period immediately after release (5.5%) as it is among 18 month survivors (3.4%), the decompositions into covariate contributions would seem to offer rather different theoretical "explanations" for these increases. In the very early period the slightly greater mean age at release of this sample and their considerably greater involvement with gangs contribute positive amounts to the increase. However, these effects are pretty much offset by a slightly lower than average rate of clearance by arrest for violent crimes reported to the police. Among the 18 month survivors none of the individual behavior or family measures are of importance. But interestingly, the lower crime clearance rate again contributes a small decrease in mean risk, now offset almost exactly by an increase attributed to the higher rate of reported violent crime.

In both time frames the variable LA County makes a direct and positive contribution. This is an effect not found in hypothetical samples in which any of the other 3 regions are over-represented[30]. While this might arguably suggest that Los Angeles County is somehow a marginally more criminogenic

---

[29]For each subject the mean inter-arrest interval was calculated as the time in crime divided by the number of prior arrests minus 1 (INCRIME/(NOARRSTS- 1)). The group mean for this statistic is 0.790 years = 288 days.

[30]North/Central is the reference category and, hence, cannot have a "direct" effect. However, the changes in mean risk "induced" by changes in the other Regional variables when considering a sample in which the representation of the North/Central region is increased by 95% are similar to those for increased Bay Area or Southern representation: mean decreases of 1.1% and 1.8% at 2 weeks and 78 weeks, respectively.

environment than the rest of the State, the explanation could well be less mysterious. The other three regions, particularly the Southern and North/Central, are made up of a mixture of a few urban areas and a number of counties that would, in comparison to Los Angeles, be considered distinctly rural. Los Angeles County, of course, is quite densely populated throughout. The small "Los Angeles County" effect might well be common to other metropolitan areas, had these been chosen as units of analysis.

4.4 Summary Comparison of Direct and Net Effects

Tables similar to Tables IV.3 might be generated for the other variables but, while each offers some opportunity for theoretical speculation, they tend to be rather repetitive. Instead, Table IV.4 gives a summary of what this sort of analysis might produce by reporting for each z_k the ratio of the direct effect on the mean risk produced by the change $d<z_k>$ itself (e.g., for three z_k the first line entries in Tables IV.3) to the net effects reported in Table IV.2. Specifically, Table IV.4 provides for each z_k the following:

$$\frac{direct(k)}{net(k)} = \frac{d<z_k>(c_k+b_kt+a_k\ln t)}{d<z_k>(c_k+b_kt+a_k\ln t)+\sum_{j\neq k} d<z_j>(c_j+b_jt+a_j\ln t)} \qquad 4.9$$

This ratio expresses the relative importance of the direct and net effects on recidivism risk. For ease of reference, the signs of the direct and net effects are also shown in Table IV.4. The blanks in Table IV.4 occur with variables that do not appear in the model for that period and, thus, are equivalent to ratio values of zero.

The following table summarizes the direct effect information for the three variables in Tables IV.3 and the net effect information in Table IV.2 that provide the entries for these three variables in Table IV.4 (at t = 2 weeks).

Variable	Direct Effect	Net Effect	Direct/Net
AGEFIRST	0.8	-5.6	-0.14
TIMEIN	0.0	-1.0	
LA	2.3	5.5	0.41

Values in Table IV.4 greater than zero indicate that the direct and net effects on risk are in the same direction, while values less than zero indicate that the direct and the sum of the induced effects (Induced = Net - Direct) are opposite in sign and that the magnitude of the induced effects is greater than the direct effect. Magnitudes between 0 and 1 suggest that, in general, the net and the (sum of the) induced effects on risk are in the same direction, while magnitudes greater than 1 mean that the direct effect on risk is larger than the net effect and thus that the direct effect is being attenuated by the (induced) changes in other variables.

TABLE IV.3c. DECOMPOSITION OF THE NET EFFECT OF d<LA> ON THE GEOMETRIC MEAN HAZARD2[1]

Variable	$\langle z_k \rangle$	$\frac{(d\langle z_k \rangle)100}{\langle z_k \rangle}$	% Change h(t,<Z>) t = 2 wks	% Change h(t,<Z>) t = 78 wks
LA	0.40	61.9	2.3	4.2

Variable	$\langle z_j \rangle$	r_{jk}	$\frac{(d\langle z_j \rangle)100}{\langle z_j \rangle}$	% Change h(t,<Z>) t = 2 wks	% Change h(t,<Z>) t = 78 wks
AGEFIRST	14.19	-0.02	-0.2	-0.0	0.1
CARLNGTH	4.14	0.09	2.6	0.6	
NOARRSTS	7.58	0.06	1.9		
PPARVIOL	1.03	0.04	2.8	0.2	0.8
PRCOMMIT	1.18	-0.15	-7.7	-1.6	-0.5
Criminal History Scores:					
VIOLENCE	1.22	0.04	2.1		
ROBBERY	0.58	0.21	15.5	0.1	
BURGLARY	1.66	0.06	2.9	0.3	-0.1
OTHPROP	1.33	0.08	4.8	0.6	0.4
GENDELQ	3.24	-0.21	-9.3	-0.0	-1.0
MWF	2.39	0.13	1.4	-0.6	
CYAVIOL	0.86	0.08	6.4	-0.1	
INFRRATE	0.82	0.02	1.8	0.2	
TIMEIN	1.13	0.12	3.3		-0.4
AGEOUT	19.45	0.14	0.6	2.6	
ALCOHOL	0.84	-0.13	-6.4	0.4	
DRUGS	1.02	0.04	1.7	-0.2	
GANG	0.47	0.36	30.1	4.2	0.1
DROPOUT	0.55	0.08	3.6	0.9	
SCHDISC	0.81	0.03	1.6		0.1
FAMSIZE	0.48	0.02	0.8	-0.1	0.1
FAMVIOL	0.40	-0.09	-8.6		
PARALCH	0.46	-0.07	-6.2	0.2	
PARCRIM	0.32	-0.01	-0.7		
SIBCRIM	0.65	0.03	1.8	0.2	-0.1
WEAKMOM	1.04	-0.02	-0.8	-0.1	-0.0
PCLRATE	0.17	-0.64	-6.0		-2.3
PCRATE	65.24	0.01	0.1	0.0	
VCLRATE	0.51	-0.57	-4.9	-4.0	
VCRATE	14.95	0.43	6.5	-0.3	2.1
Region:					
BAYAREA	0.20	-0.41	-40.4	-0.2	
SONOTLA	0.19	-0.39	-40.4		
Net Effect				**5.5**	**3.4**

[1] This is a sample in which the regional representation is made up of LA County: 64%, Bay Area: 12 %, South: 11%, North/Central: 13% instead of 40%, 29%, 19% and 21%, respectively. Blanks correspond to variables that do not appear explicitly in the model for that period (see Table III.1).

To frame this analysis in terms used earlier is this chapter, we can consider the ratio of direct to net effects on risk defined in equation 4.9 as a comparison of two hypothetical populations who differ from our observed population as follows: The hypothetical "direct effect" population has mean covariate values $<Z>$ that differ from our observed population only on variable k; specifically, the mean of this variable differs from that of the observed population by $d<z_k> = s_k/2$, where as before $<z_k>$ and s_k are the mean and standard deviation of the kth variable in the observed population.[31] The hypothetical "net effect" population is the population discussed in section 4.2, which differs from our observed population by $d<z_k>$ on k as well as by the correlated values $d<z_j>$ on the remaining variables in the model.

The distinction between the direct and net effects is illustrated clearly by the variable number of prior arrests (NOARRSTS). In model construction this variable was not found to have consistent risk discriminating power independent of that ascribed to other variables (i.e., it was dropped from the model). Taken literally, then, the model considers all "hypothetical" pairs differing only in their number of prior arrests to be equal risks -- the direct effect. But prior arrests are strongly associated with other criminal history variables and, as shown in Table IV.2, there is a quite large and persistent difference in net risk associated with a mean population difference of 2.3 arrests, ranging from a percentage increase in geometric mean hazard of from 16.3 percent (at t = 2 weeks) to 22.1 percent (at t = 156 weeks).

The seriousness of the charge at current commitment (MWF) offers an example in which increased values tend to decrease risk -- at least in the early period after release. A hypothetical population that differs from the observed only in an increased seriousness of the offense of current commitment would, according to this model, pose a lower average risk than the observed population (the direct effect). But increases in MWF also tend to be associated with lower values of other variables (mainly criminal histories) that the model considers important in estimating risk levels. Thus the direct effect of MWF alone underestimates the amount of mean risk reduction associated with a typical sub-population committed on more serious charges.

Ratios in Table IV.4 greater than +1 again indicate risk differences in the same direction for the hypothetical "direct effect" and "net effect" populations. But, in this case, increases in variables typically correlated with z_k tend to act in the opposite direction from z_k itself. Thus, gang involvement, for example, is considered strongly related to risk in the period immediately after release (direct/net for GANG = 1.24 at t = 2 weeks). But gang membership also tends to be more characteristic of subjects who are both younger at release and come from counties with lower than average violent crime clearance rates, attenuating somewhat the effect ascribed to the variable GANG alone.

For a few of the variables in Table IV.4 the ratios are negative, indicating opposite directions for the direct and net effects. DRUGS in the period very early after release is an example. In this case, the direct effect is negative and the net effect is positive (direct/net = -2.13). This variable, based pretty much on self-reports, describes drug use during a period that pre-dates the "crack epidemic." As shown in Table IV.2, the variable DRUGS does not seem to have much discriminatory power. However, as we shall see in the next few paragraphs, this variable offers an interesting illustration of some of the complexity involved in trying to interpret the model's handling of interactions between variables.

[31] A difference of precisely this value could, of course, only occur with continuous variables.

TABLE IV.4. RELATIVE IMPACT OF DIRECT AND NET EFFECTS OF $d<z_k>$ ON GEOMETRIC MEAN HAZARD RATES[1]

Variable	$(\text{Direct Effect})_k$ / $(\text{Net Effect})_k$ Weeks Since Release 2	78
AGEFIRST	-0.14 (+/-)	0.43 (-/-)
INCRIME	0.40 (+/+)	
NOARRSTS		
PPARVIOL	0.45 (+/+)	0.84 (+/+)
PRCOMMIT	0.75 (+/+)	0.36 (+/+)
Criminal History Score:		
VIOLENCE		
ROBBERY	-0.40 (+/-)	
BURGLARY	0.82 (+/+)	-1.08 (-/+)
OTHPROP	0.73 (+/+)	0.54 (+/+)
GENDELQ	0.00 (+/+)	0.37 (+/+)
MWF	0.82 (-/-)	
CYAVIOL	-0.31 (-/+)	
INFRRATE	0.78 (+/+)	
TIMEIN		0.49 (-/-)
AGEOUT	1.21 (+/+)	
ALCOHOL	-10.07 (-/+)	
DRUGS	-2.13 (-/+)	
GANG	1.24 (+/+)	0.07 (+/+)
DROPOUT	0.78 (+/+)	
SCHDISC		0.39 (+/+)
FAMSIZE	2.26 (-/-)	1.22 (+/+)
FAMVIOL		
PARALCH	1.20 (-/-)	
PARCRIM		
SIBCRIM	0.76 (+/+)	-2.91 (-/+)
WEAKMOM	1.27 (+/+)	0.43 (+/+)
PCLRATE		2.83 (+/+)
PCRATE	-0.81 (+/-)	
VCLRATE	6.03 (+/+)	
VCRATE	1.92 (-/-)	1.18 (+/+)
Region:		
LA	0.41 (+/+)	1.24 (+/+)
BAYAREA	-0.12 (+/-)	
SONOTLA		

[1] Entries are the ratio of the changes in geometric mean hazard due to the direct and net effects of a change in $<z_k>$, see equation 4.9. Blanks indicate that the variable does not enter the model directly, implying a ratio of zero.

Table IV.2 shows that an increase in the mean of the Drug Use score from 1.02 to 1.42 would produce a rather modest 2.2% increase in risk in the first few weeks after release. But from Table IV.4 we see that the model associates a decrease in risk with this variable if the comparison is between two hypothetical populations differing only on drug use. The typical drug user in this population tends to have a longer criminal history, with more burglary and general delinquency charges but fewer charges for the more serious crimes of violence and robbery. He is more likely to have a prior commitment, to be somewhat older when released, to have dropped out of school and to have a record of gang membership. These additional risk factors associated with drug use outweigh the negative influence the model attributes explicitly to this variable alone.

But the question remains: Why should the model assign a higher risk to the member of a hypothetical pair who is recorded as not being a drug user? Not only would this seem to contradict past research and current theory; but in fact observed failure rates among drug users in this population are higher than those of non-drug users over a number of different time frames that were examined. The analytic answer to this seeming paradox is contained in the data shown in Table IV.5.

The first three columns of data divide the 1949 subjects into groups with different drug use scores.[32] The last column combines the two drug-using groups. The four variables whose means are given here are the ones making the largest induced contribution to the net effect for Drugs in Table IV.2 -- positive for AGEOUT, DROPOUT and GANG; negative for ALCOHOL. It should be noted that all of them increase monotonically with increasing severity of a drug problem.

What is remarkable in this table is that in the period immediately after release (illustrated here as 12 weeks) the observed failure rate for the group with DRUGS = 2 is lower than that for the Drugs = 1 group. Is this difference "real?" Under the null hypothesis a simple chi-squared test estimates the probability of such a difference arising by chance at 0.100. Although this is hardly enough evidence to build a theory on, the data of Table IV.5 would at least seem to explain the anomaly of the model's assignment of lower early risk to higher values of drug use. In effect it may be considered as a correction factor that is important only in the very early period. By 36 weeks, the coefficient for DRUGS is virtually equal to 1. But without this factor the model would, on the basis of other covariates, assign a substantially higher initial risk to the typical subject seriously involved with drugs -- contrary to what is observed in this population.

[32]The categorical variable DRUGS was scored, based on official records, as 0 if there was no evidence of drug use, 1 if there was evidence of minor drug use, and 2 if there was evidence of serious or major drug use.

TABLE IV.5. DRUG USE AND EARLY RECIDIVISM

	Value on Variable DRUGS[1]			
	0	**1**	**2**	**> 0**
Variable	Means			
AGEOUT	19.2	19.4	19.8	19.6
DROPOUT	0.47	0.55	0.61	0.58
ALCOHOL	0.42	0.86	1.22	1.03
GANG	0.39	0.45	0.57	0.51
Period Since Release	Failure Rate			
First 12 Weeks	0.17	0.21	0.17	0.19
First 36 Weeks	0.48	0.50	0.52	0.51
N	611	695	643	1338

[1] The variables DRUGS, ALCOHOL and GANG were coded 0, 1, or 2 if the official record indicate no, minor, or major evidence, respectively, of involvement.

CHAPTER V
CLASSIFICATION FOR RISK:
The Relative Improvement Over Chance and Related Statistics.

5.1 Introduction

For many purposes--especially for exploration of policy and program options--an individually estimated failure probability is too fine-grained to be useful as a classification system. What is wanted is a method for sorting subjects into a few classes (perhaps high, medium and low risks) and an analysis that describes some of the likely consequences attendant on the adoption of any particular classification. In this chapter we regard the model-assigned failure probabilities as a basis of such a classification but only in the sense of providing a reasonable rank-ordering of subjects by rearrest risk.

It is clear from the analyses of the Chapter III that the model does not identify any group of subjects that could be defined to be truly low risk in terms of their probability of surviving arrest-free for at least three years. As used here the term "low risk" must be regarded as something of a euphemism, defining a class of subjects whose risk is "low" only in comparison to a "high risk" class. Nevertheless, we shall throughout make use of terms that have become part of the lingua franca of classification studies. In particular, the analyses will focus on "error" rates, estimating the numbers of "false positives" and "false negatives" resulting from any particular separation of the population of interest into high and low risk groups.

The degree to which such terms should carry the weight of their common, lexical definitions depends in part on what one expects from a risk classification system. For example, if consideration is being given to the adoption of very expensive or perhaps draconian measures to reduce recidivism, one would like the target population to be made up of subjects whose rearrest is virtually certain under current control policies. Or at the other extreme, concerns about public safety may be paramount if for some reason a quite substantial relaxation of the current level of control for some subset of the population is being considered.

Given a probabilistic view of the recidivism process, one would not consider it an "error" to find, say, a 35% success rate in a subset of the population defined as "high risk" because their mean failure probability is estimated to be 0.65. It would certainly be a mistake, however, to develop control policies that simply ignore expected survival or failure rates ("error" rates) in groups categorized as high or low risk, respectively.

The next section of this chapter briefly describes the Relative Improvement Over Chance (RIOC) statistic -- a generally applicable measure of a classification system's power to discriminate between high and low risk groups. Section 5.3 contains empirical results, assuming that the model is used as a basis for classifying the entire study population as either "high" or "low" risk individuals. The final section illustrates the use of a probability model for a triage classification in which allowance is made for a "medium risk" category.

5.2 The Relative Improvement Over Chance (RIOC)[33]

A number of different statistics have been proposed in the literature as overall measures of a classification system's predictive accuracy (see Gottfredson and Gottfredson, 1985, 1986). Here, the analysis is based in great part on the Relative Improvement Over Chance (RIOC) (Loeber and Dishion, 1983; Farrington and Loeber, 1989). This is defined as the improvement in the number of

[33]See Appendix B for algebraic derivations of relations appearing in this section.

correct predictions of failure (or success) over what would be expected if subjects classified as high risk were chosen randomly from the population. It is normalized by expressing the statistic as a fraction of the maximum number of correct failure (or success) predictions that could possibly be made, given that a specified number of subjects E are designated high risk and there are a total of F observed failures in population of size N:

$$RIOC = \frac{\left(A_{obs} - \frac{EF}{N}\right)}{\left(A_{max} - \frac{EF}{N}\right)} \qquad 5.1$$

Here A_{obs} is the observed number of high risk failures ("true positives") and

$$(A_{max} - \frac{EF}{N}) = \begin{cases} (E - \frac{EF}{N}) & E < F \\ (F - \frac{EF}{N}) & F < E \end{cases} \qquad 5.2$$

We choose to focus attention on reductions in error rates rather than improvements in correct predictions. It will, therefore, be convenient to express the RIOC in a somewhat different form.

We define B to be the observed number of "high risk" successes ("false positives") and C the observed number of "low risk" failures ("false negatives"), with $\langle B \rangle$ and $\langle C \rangle$ the corresponding numbers that would be expected if E subjects were randomly selected to form the high risk group:

$$\langle B \rangle = E\frac{(N-F)}{N} \qquad 5.3$$

and

$$\langle C \rangle = (N-E)\frac{F}{N} \qquad 5.4$$

After a bit of algebraic manipulation the RIOC can then be written

$$RIOC = \begin{cases} 1 - \frac{B}{\langle B \rangle} & E < F \\ 1 - \frac{C}{\langle C \rangle} & E > F \end{cases} \qquad 5.5$$

In this form the RIOC has a straightforward interpretation in terms of the choice of the selection ratio, E/N, and the implied objectives of the 2x2 classification. A relatively low value might be chosen for the selection ratio (E < F) if it is important to ensure that the high risk group contains only those subjects about whose failure the model purports to be most certain. The RIOC then is a direct measure of the power of the model's classification over that of pure chance in terms of the fractional reduction of "false positives." A RIOC of 0.60, for instance, means an expected 60% reduction[34] in the "false positive" rate. A parallel argument obviously holds if the dominating concern is for public safety and the model is used to make a conservative determination of subjects to be classified as low risk (E > F). The RIOC, of course, is then a measure of the model's expected power to reduce the "false negative" rate.

5.3. Risk Classification Results

Figure V.1 shows the results obtained in predicting rearrest within 3 years with an increasing fraction E/N of the population designated as "high risk." Here subjects were rank-ordered from high to low by their model-assigned probability of rearrest in 3 years[35]. The high risk group is defined to be the population fraction shown along the x axis with the selection starting from the top of the ordered risk listing. Each increment of 0.05 represents the inclusion in the high risk category of the next 97 subjects in the list[36].

The population 3 year failure rate is 0.853, which is, of course, the failure rate we should expect a priori in any randomly chosen sub-group. The observed failure rate among the high risk group starts at 1. After some initial instability, this rate decreases monotonically to 0.874 at x = 0.95.[37]

The mirror image of the high risk failure rate, the high risk survival rate, is also shown since this is the "false positive" rate. With 5% of the population rated high risk, the failure rate for the low risk group, the "false negative" rate, is 0.846. This decreases monotonically but still has the quite high value of 0.469 with only 5% of the population considered low risk.

The remaining curve in Figure V.1 is the RIOC curve. This curve initially expresses, as a fraction of the population survival rate (0.147), the difference between the high risk survival rate expected under random assignment and the rate actually observed. That is, it is initially the fractional decrease in the "false positive" rate. This curve again shows some initial instability but then decreases monotonically to 0.329 when 85% of the population is classified high risk. Thereafter it increases again; but, since the number of high risk subjects (E) is now greater than the total number of failures in the population (F), the RIOC is expressing the fractional decrease in the "false negative" rate.

It might be noted that the overall shape of these curves offers some assurance that the model-based ordering of individual risks is in general agreement with the observed group outcomes. In particular, if the model's rank ordering of rearrest risk is indeed valid, the RIOC must pass through a minimum when E = F. As the high risk group is being expanded, it is acquiring subjects ranked lower on the risk scale, taking them from the top-ranked of the low risk subjects. An increase in the fraction classified as high risk thus represents, according to the model, a continual decrease in the high risk

---

[34]We emphasize again that this is a percent of the false positive rate anticipated under a random assignment with a given number E to be ranked high risk.

[35]Since the classifications considered in this chapter depend only on a rank-ordering of probabilities, there are a number of reasonable choices that might be made of the particular model-assigned probability to use. One could, for example, use 3 year failure probabilities throughout, whether or not the classification is for risk of failure within 3 years. It was generally found, however, that better results were obtained (as measured by comparison of RIOCs) when the probability corresponded to the event for which risk was being assessed -- failure within 36 weeks, perhaps, or failure within 3 years conditioned on some period of survival. Incidentally, this also shows that as time goes on subjects' risks do not all maintain constant rank-order positions relative to one another.

[36]Or 98 if Nx is an integer.

[37]For the first four points the fractions of high risk survivors are 0/97, 5/194, 7/292 and 7/389.

group's homogeneity with respect to the probability of rearrest but an increased homogeneity of the low risk group. As the size of the high risk group passes through the point where E = F, the RIOC by definition shifts its target from the high risk to the low risk group-- that is from "false positives" to "false negatives." Thus, we should expect the RIOC, which is in effect a measure of outcome homogeneity of the target group, to behave much as it does in Figure IV.1.

Even if we accept the validity of the model's rank-ordering of risk, these results may not be particularly useful for policy development. While they indicate the possibility of isolating a high risk group whose failure within 3 years can be predicted with considerable confidence, they also show that at best this model could identify a small low risk group whose 3 year survival is about as certain as the result of a coin toss. For practical purposes it may be more useful to consider a risk classification that proceeds in stages, first considering risk of early failure and subsequently reclassifying those subjects who have actually survived arrest-free for some given length of time.

In Chapter III it was suggested that the model seems to be doing a reasonably good job of assigning probabilities of failure within the first 36 weeks. The identification of a group of "chronic" failures could have important implications for differential treatment. At the very least parole authorities could expect to have a longer period of time to work with "non-chronics" and possibly a greater chance of achieving positive, rehabilitative results. But such a program would require a "yes" or "no" decision about whether a subject is a member of the "chronic" group and the model delivers only a probability of membership.

Figure V.2 shows the results of a classification by risk of failure within 36 weeks, again as a function of the fraction of the total population to be identified as high risk. The population failure rate here is 0.502 and the graphs are almost symmetric about the 50% high risk line. If the principal concern were to avoid "false positives," the RIOC values suggest that the classification could produce reasonably acceptable results as long as the high risk fraction were kept fairly small -- say 10 or 15 percent. Of course, this entails a false negative rate of the order of 0.45. But policy based on this classification might still be found defensible in terms of its improvement over a policy that implicitly assumes that all subjects pose the same risk.

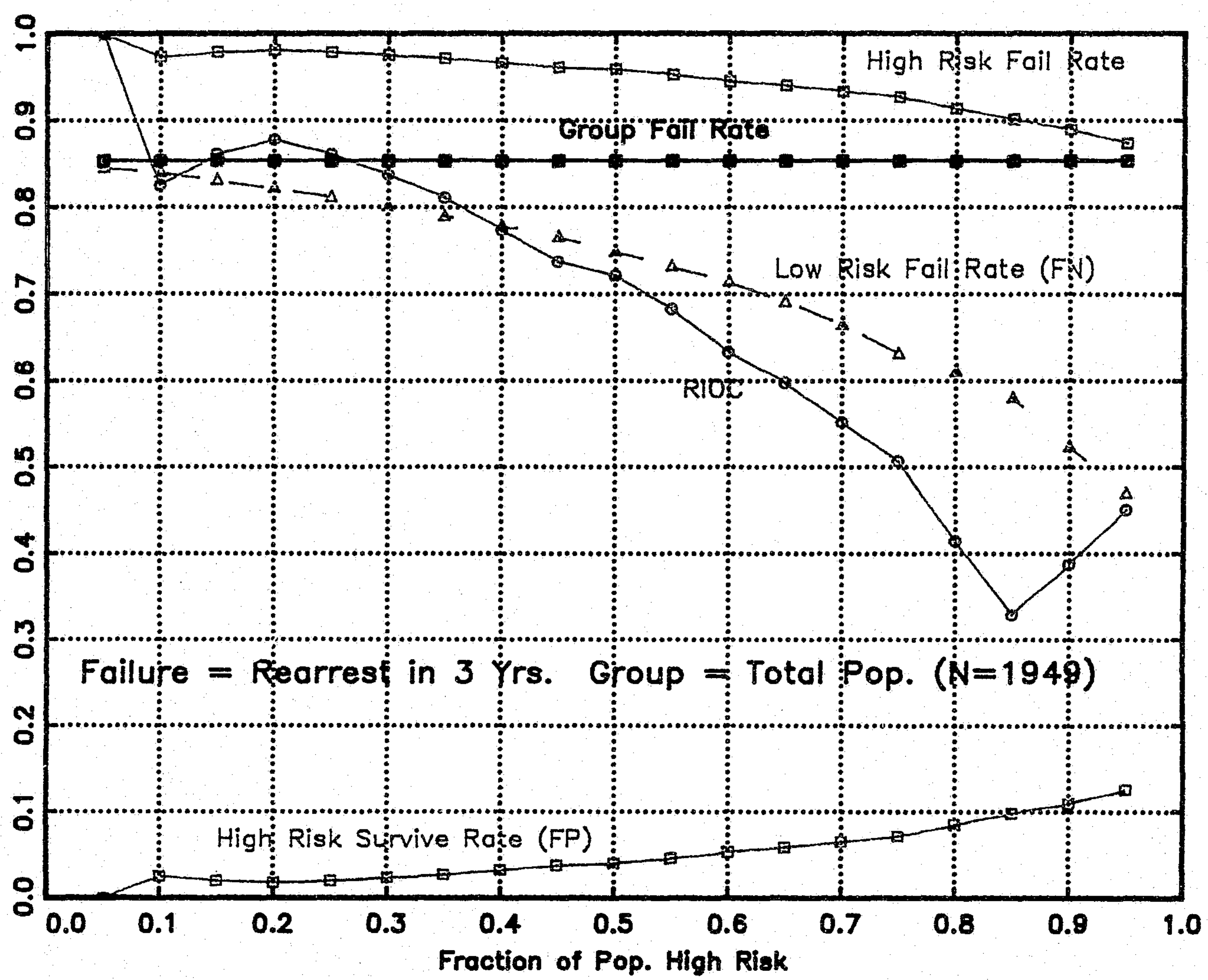
At the other extreme, if the concern is for public safety and the avoidance of false negatives, a relatively safe 10 or 15 percent of the population might be classified as the "non-chronic" group but this again means accepting an error rate of about 0.45 in the designation of the "chronics."

This naturally raises a question of the extent to which a low 36 week risk group identified here is in fact made up of 3 year successes. Of the lowest ranked 195 subjects, 39 were rearrested within the first 36 weeks (Figure V.2's "false negatives" when the fraction rated high risk is 0.90). Only 80 survived arrest-free for three years. Certainly the 0.41 three year success rate of this group compares favorably with 0.16 success rate of the remainder of the population. But a treatment policy based on an initial assessment of the risk of early failure must realistically take into account the fact that the number of subsequent failures can be quite high -- even among those deemed least likely to fail early.

As time goes on the identification of probable successes is aided pragmatically by the simple fact that the population of survivors is shrinking. Figures V.3a and V.3b examine the results of a reclassification of the 970 subjects surviving at least 36 weeks and of the 413 surviving at least 2 years[*]. In both cases, failure is defined as rearrest within 3 years after release and risk is based on the rank-ordering of the corresponding conditional failure probabilities.

[*]Note that increments of 0.05 in the fraction ranked high risk here correspond to transfers from low to high risk categories of 48 (or 49) subjects in Figure V.3a and 20 (or 21) in Figure V.3b.

Figure V.1



High Risk Fail Rate
Group Fail Rate
Low Risk Fail Rate (FN)
RIOC
Failure = Rearrest in 3 Yrs. Group = Total Pop. (N=1949)
High Risk Survive Rate (FP)
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
Fraction of Pop. High Risk

Figure V.2

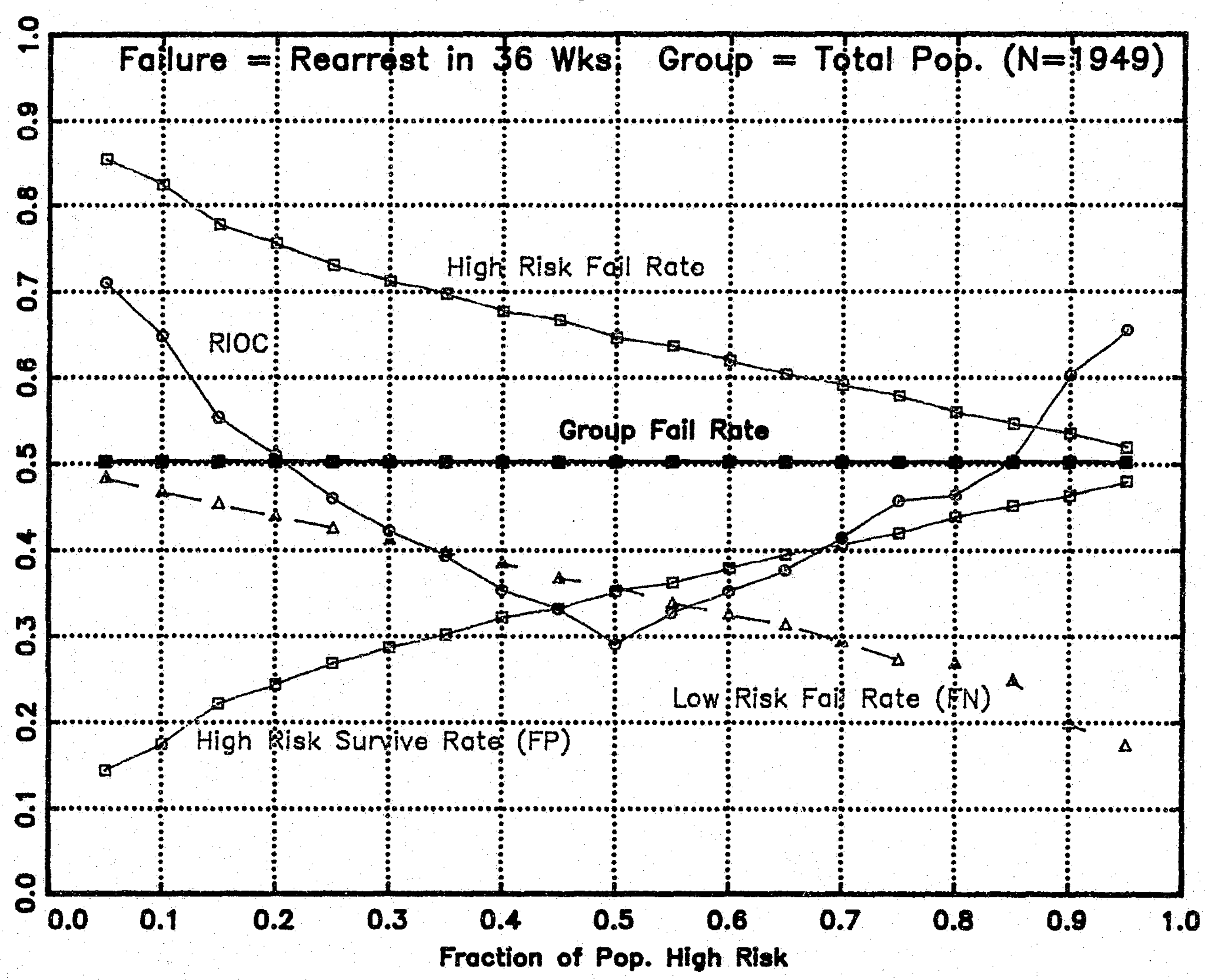

Failure = Rearrest in 36 Wks. Group = Total Pop. (N=1949)
High Risk Fail Rate
RIOC
Group Fail Rate
Low Risk Fail Rate (FN)
High Risk Survive Rate (FP)
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
Fraction of Pop. High Risk

Figure V.3a

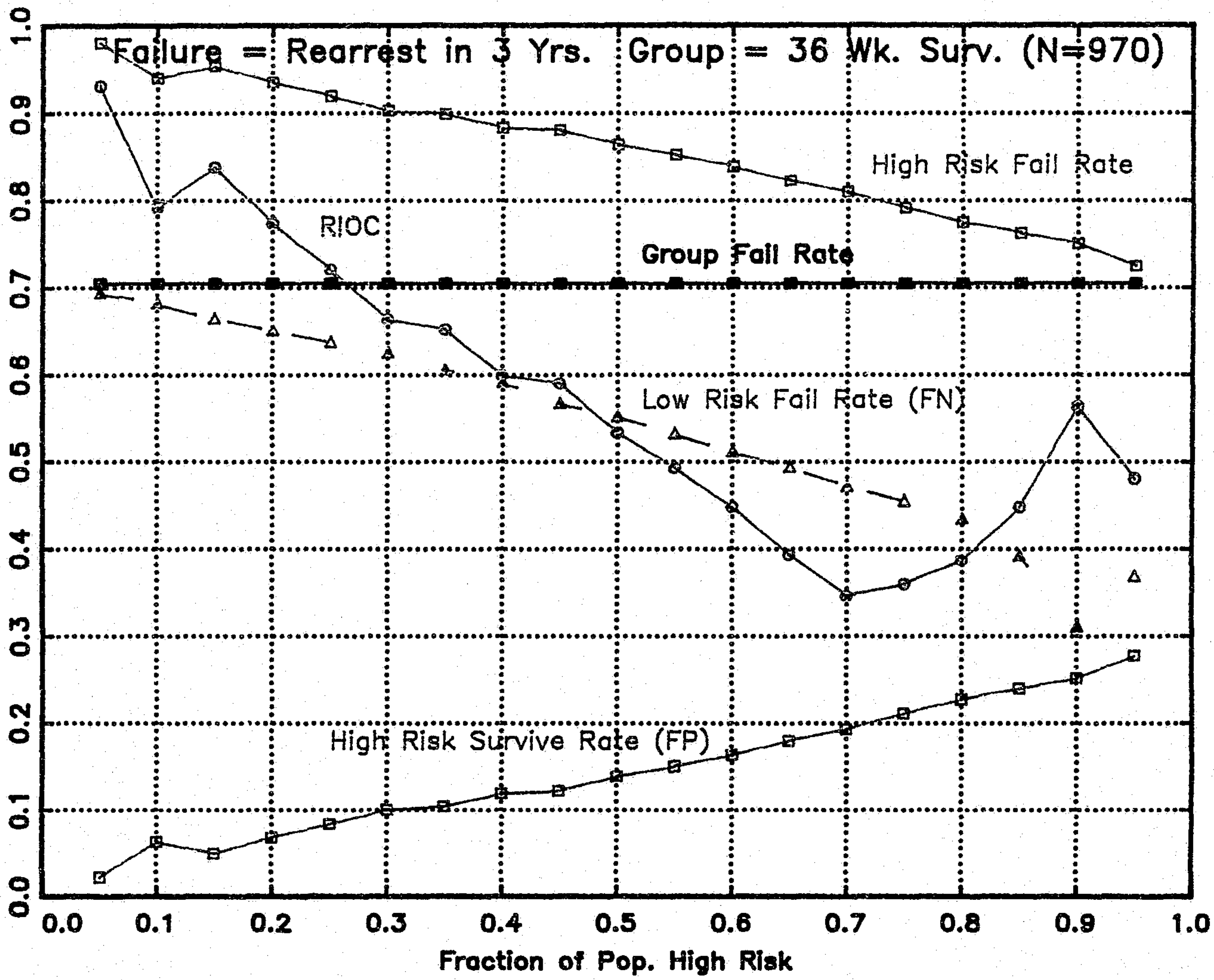
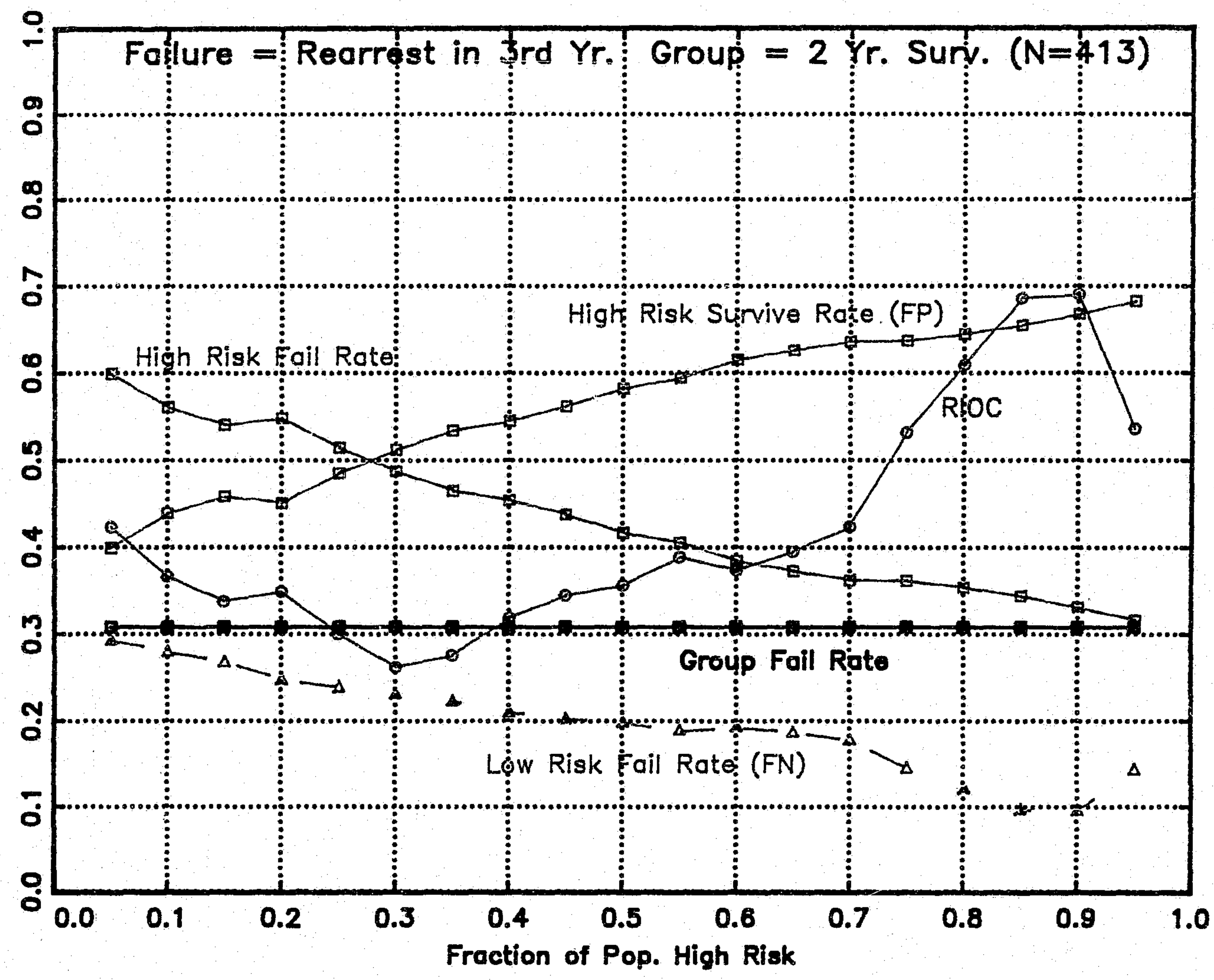

Failure = Rearrest in 3 Yrs. Group = 36 Wk. Surv. (N=970)
High Risk Fail Rate
RIOC
Group Fail Rate
Low Risk Fail Rate (FN)
High Risk Survive Rate (FP)
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
Fraction of Pop. High Risk

Figure V.3b


Failure = Rearrest in 3rd Yr. Group = 2 Yr. Surv. (N=413)
High Risk Survive Rate (FP)
High Risk Fail Rate
RIOC
Group Fail Rate
Low Risk Fail Rate (FN)
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
Fraction of Pop. High Risk

The group failure rate among 36 week survivors (Figure V.3a) is 0.705. As a predictor the classification would again be most accurate in selecting a high risk sub-group. Among 2 year survivors (Figure V.3b) the failure rate during the third year at risk was 0.308. Here the classification favors a rather cautious prediction of success for a small group of subjects at the low end of the distribution-- as evidenced by the relatively large values for the RIOC and the relatively small values for the low risk failure rate when more than 80 percent of the population is assigned to the high-risk group. These results should be compared with those of the initial, 3 year forecast shown in Figure V.1.

Of course, the results being considered here differ from the initial results not only in the loss of earlier failures from the population being classified but also in the shorter time period during which failure as defined can occur. Does the gradual emergence of a small sub-group of subjects that might be considered relatively "safe" risks actually reflect a lower propensity for rearrest among these longer term survivors? A partial answer to that question can be obtained from the data used to generate Figure III.2, in which the numbers of observed failures are shown by 4 week intervals. Dividing these numbers by the population surviving to the beginning of each interval, we obtain a gross measure of the surviving population's mean risk as a function of time. In fact, this empirical hazard rate increases rapidly over the first few months after release. Between weeks 8 and 12, 8.3% of the population surviving at least 8 weeks were rearrested; between weeks 12 and 16, 9.0% of the 12 week survivors failed; and from weeks 12 to 16, 9.6% of those surviving 12 weeks failed. Thereafter, these rates drop slowly but steadily (with some not unexpected randomness) to a 4 week failure rate of 2.1% of survivors at week 152.

5.4. "Triage" Classification

Up to this point, it has been assumed that the policy considerations motivating a risk classification are dominated by concerns about only one error type: either the false positive or the false negative rate. All subjects are to be classified and everyone not defined as "high risk" is "low risk" by default. For the design of some types of programs, however, it might be more sensible to use risk classification as a system of "triage," specifying both "high" and "low" risk groups, with "medium" risk as a default category. For example, this would seem to be the case if because of resource constraints increased attention to a high risk group necessarily implies decreased supervision of low risk subjects, with treatment of "medium" risks presumably remaining unchanged.

As an illustration we consider such a system of triage applied to the prediction of failure within the first 36 weeks after release. Figure V.4a shows the results of increasing the size of the low risk population under a policy that is required to be quite conservative about defining subjects as high risks -- here only the top 5% of the total population. Figure V.4b gives the complementary results if the dominant concern were avoidance of "false negatives".

It is important to note that the number of subjects (N) and the number of failures (F) are combined counts from the two groups actually classified -- not from the whole population. "Chance" in the RIOC definition therefore refers to a random selection of E "high risk" subjects from the classified pool. The RIOC may be thought of as measuring the improvement in predictive power achieved by considering the high and low risk groups as separately homogeneous with respect to outcome compared to what might be expected were they considered as a single, homogenous sub-population.

In Figure V.4a, the number of high risk subjects is held constant (E = 97) and at all data points this number is less than the combined number of failures. Consequently, the RIOC curve describes the fractional decrease in the false positive rate. As the size of the group defined as low risk grows, the combined high plus low risk groups' failure rate decreases until the low risk class contains the lowest ranked 25% of the population. Continuing to expand the size of the classifed group beyond this point means including ever more dubious cases in the low risk category and thereafter the overall group failure rate increases monotonically.[39] In Figure V.4b it is the low risk group that is being held constant (N-E = 98) and, except for the first point, the RIOC curve measures the fractional reduction in false negatives.

In both cases the RIOC values are relatively high and show little variation over the range of x. This is a reflection of the very conservative selection of the target populations (high risk in Figure V.4a; low risk in V.4b) and the relative homogeneity of these sub-populations with respect to outcome.

5.5 Concluding Remarks

This chapter has been somewhat abstract and it might be useful to conclude with a very brief discussion of the implications these RIOC analyses might hold for policy and theory.

The results here indicate that one can array the population according to subjects' relative risk of rearrest and be reasonably assured that the ordering will be consistent with comparative failure rates of defined sub-populations. But the model from which these results are derived does not provide decision makers with a crystal ball that can infallibly look into the future. For most groups defined from the population the absolute level of risk leaves ample room for "chance" to play a role in determining the outcome for individual subjects.

In a sense it is this element of chance that is the target of policies and programs aimed at reducing recidivism. The practical utility of the information conveyed by a risk-ordered classification thus depends on whether innovations being considered could take advantage of inter-group differences in the degree to which subjects' failure or success is a gamble. If 85% of the total population can be expected to fail within 3 years under current policies, is it worthwhile to be able to identify sub-populations currently having 95% or 60% expected failure rates? The answer, of course, depends on estimates of changes in net costs and failure rates anticipated as resulting from a differential change in group treatments.

From a theoretical perspective it is, perhaps, disappointing that the model is not more efficient in early identification of a group that is truly low risk among the 286 subjects surviving for three years. Even among those subjects not rearrested for two years after release, chance would seem to be a significant factor in determining third year success or failure.

It is possible, of course, that a truly low risk group does not exist within this population -- that the model captures the essentials of risk discrimination between subjects and everything else is in fact just chance. But it is also possible that we are faced with a case of unobserved heterogeneity. There may be unmeasured variables that would help to distinguish a sub-group of truly low risk subjects among those classified by the model only as relatively low risk.

It might be remarked that all the variables used here tend in theory to be symptomatic of a social pathology. The best the model can do in terms of identifying a "safe" risk is to note a relative absence of these symptoms. Of course, these are the variables that criminological studies have found associated with recidivism. But in a population such as the one studied here, it is the long term success that is the rare event.

---

[39]Algebraically, if F and N denote total group failures and group size at any given x value and dF is the number of failures in the next increment dN in group size, then for $x < 0.25$ in Figure IV.4a, $dF/dN < F/N$. For larger x the direction of the inequality is reversed. The RIOC also passes through its maximum (0.771) at this point. Similar remarks apply to the curves of Figure V.4b, where the group failure rate minimum and the RIOC maximum (0.728) occur with the top 20% of the population classified high risk.

Figure V.4a

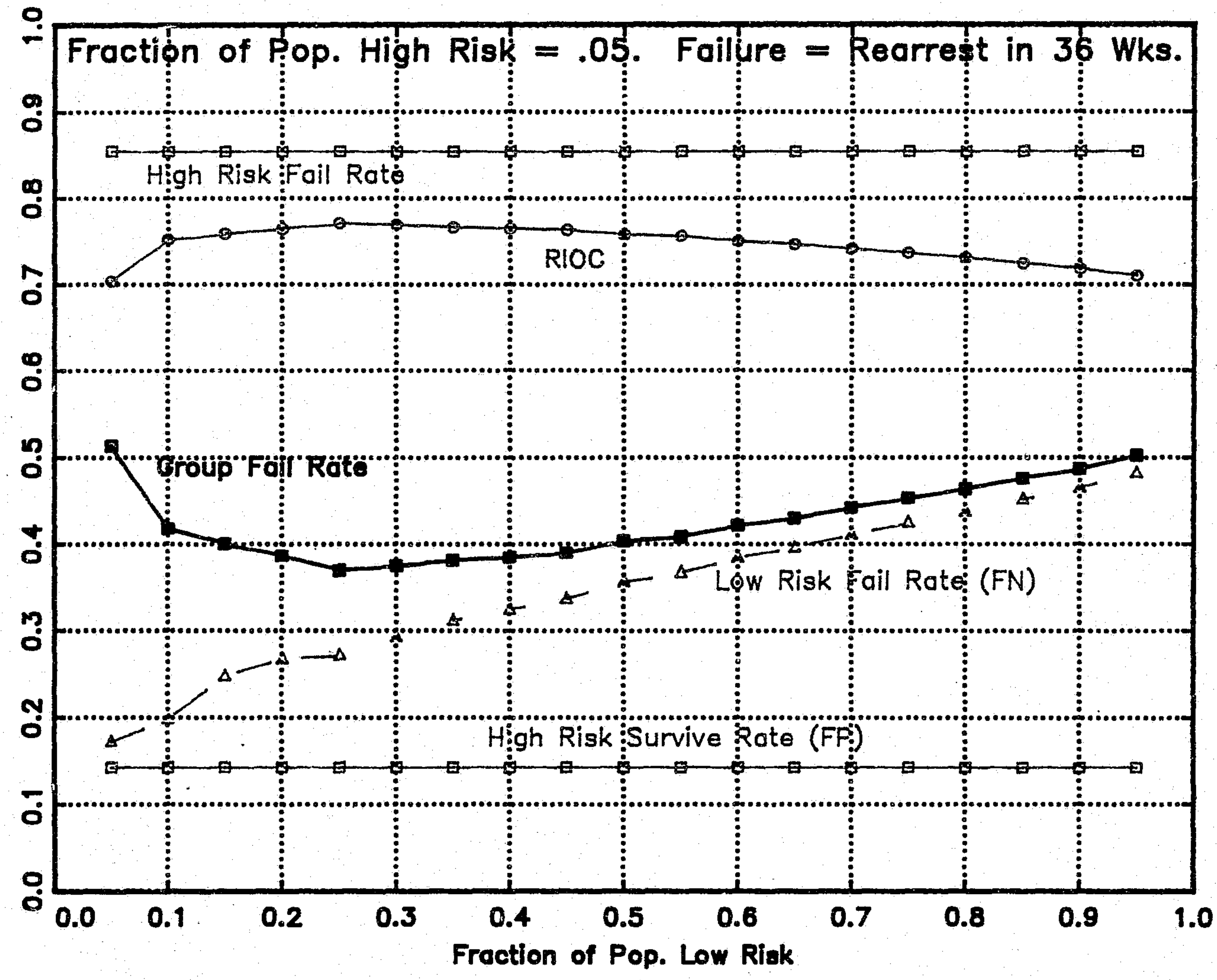
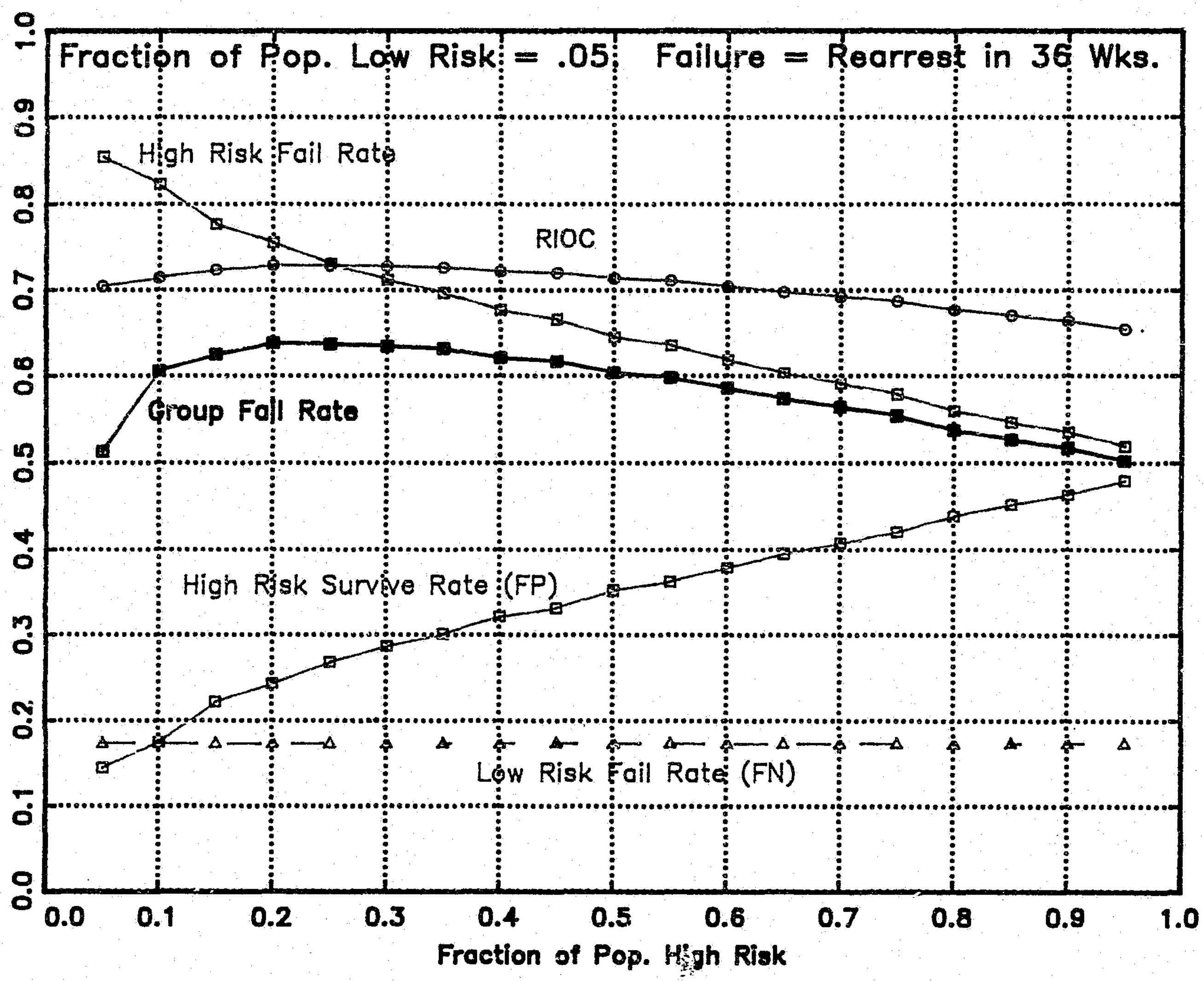

Fraction of Pop. High Risk = .05. Failure = Rearrest in 36 Wks.
High Risk Fail Rate
RIOC
Group Fail Rate
Low Risk Fail Rate (FN)
High Risk Survive Rate (FP)
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
Fraction of Pop. Low Risk

Figure V.4b



Fraction of Pop. Low Risk = .05; Failure = Rearrest in 36 Wks.
High Risk Fail Rate
RIOC
Group Fail Rate
High Risk Survive Rate (FP)
Low Risk Fail Rate (FN)
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
Fraction of Pop. High Risk

The predictive efficiency of models such as this might therefore be enhanced if a few reliably measured variables were included that are associated in theory with good social adjustment -- measures such as those developed for assessing pre-trial risk, perhaps. Such variables would not have to be particularly powerful as bivariate predictors of success or failure. They could perform a very important function if they only added greater discriminatory power at the low end of the risk scale.

CHAPTER VI
MODEL APPLICATIONS

6.1. Potential Benefits

The three preceding chapters of this paper have been concerned primarily with technical issues: the development of a model for the statistical prediction of rearrest and the application of some simple analytic tests to examine how the model operates and how well it performs. We now want to turn to issues of more immediate practical utility. How might parole officers and administrators make use of analytic results from a model like this, given that it adequately reflects conditions prevailing under current policy?

Four subjects were selected from the population for purposes of illustration. Their data base descriptions are given in Table VI.1. Subjects 1235 and 64 were chosen because they had, respectively, the highest (0.52) and lowest ($< 10^{-12}$) modeled probabilities of surviving for three years after release. Subject Number 615 is taken from the middle of the probability distribution and represents a risk level that is typical of much of the population, with a 3 year survival probability of 0.09. Finally, Subject 292 (S(3 yrs) = 0.04) was selected because, among 3 year survivors, his hazard function shows the greatest variation over time.

Two kinds of model-generated output are shown in Figures VI.1 and VI.2. Figure VI.1 reflects the model's changing assessment of the rearrest risk each of these subjects poses. It is essentially a sort of "macro-hazard" function, with each point on a curve specifying the probability that the subject will be rearrested during the next six months under the supposition that he has not yet been rearrested at that point.

The median risk case, 615, was 14 years old when first arrested. He seems to have been seriously delinquent with a record of drug and alcohol abuse, gang involvement and school problems. But rearrests were fairly sporadic with a mean interval of about 2.25 years. He also would seem to have something of a penchant for robbery, which perhaps accounts for his 1 prior commitment. When released from the 2 or 3 months of his current commitment, he was about 2 years older than the average of this population. His six month ahead forecast is quite flat, rising from an initial rearrest probability of 0.30 to about 0.37 at 36 weeks and then slowly decling to 0.24 at 2 1/2 years. He survived arrest-free for almost a year but was picked up on day 349 and charged with a serious property crime.

Case 292 began his arrest career much earlier than 615 and subsequently had considerably more frequent encounters with the criminal justice system. On average he was rearrested about every 7 months. But, with 2 prior commitments, he was presumably not at risk of arrest for some significant fraction of his 9 year career. Although not a stranger to violence, his preference would seem to have been burglary and serious property crimes along with assorted, more minor offenses. Interestingly, except for some gang involvement, his record does not indicate a pattern of juvenile delinquency as measured by substance abuse or school problems. His initial 6 month failure probability is quite high (0.64). After a slight rise in the first few weeks, it begins to drop off quite rapidly to 0.35 by week 36. After that the decrease is quite slow and, in fact in this later period, the risk the model assigns to 292 is virtually the same as that for 615. Despite 292's more extensive arrest history, he has no record of previous parole violations -- unlike 615. As shown in Table IV.2, this variable is considered by the model as a very strong indicator of current risk among the population of subjects who have survived for an extended period of time. For over 4 years after release from his 2.3 year current commitment subject 292 has no record of an arrest.

The model gives almost no hope of post-release success to a subject like 64. He was almost 13 when first arrested but in the next 5 years accumulated an astonishing record of a rearrest about every 3 months. These were predominantly arrests on relatively minor general delinquency charges and did not result in a commitment prior to the current one. Apparently, however, he violated conditions of prior releases with some regularity. He seems to have had a serious record of juvenile delinquency (substance abuse, gang involvement, school problems). And, finally, after a current commitment of 9 or 10 months, he was still quite young at the time of his release. His conditional probability of rearrest within the next 6 months begins with a high value of 0.84 and increases to 0.999 at 2 1/2 years. His rearrest on day 314 was again for a general delinquency offense.

Case 1235 had no criminal record prior to his arrest on a violent felony charge when he was 19 years old. He had previously dropped out of school and showed evidence of serious problems with drugs and alcohol. He is recorded as coming from a family with problems, including family violence. But in the Southern California county in which he was convicted, the local rates of both property and violent crime were well below the average for this population. He was also about 2 years older than the average when released after serving about 15 months. The model's six month ahead forecast for this subjects starts with the relatively low value 0.16, decreasing thereafter to 0.03 after 2 1/2 years. There is no record of his being rearrested during more than 3 1/2 years of follow-up.

The six month forecast period of Figure VI.1 was, of course, arbitarily chosen. Graphs like these could be generated for each subject with the forecast period being any given interval of interest. It might be noted that, as conditional probabilities, such results could be used to estimate quantities of possible administrative interest such as the expected fraction of the surviving population that will fail during the next interval.

Figure VI.2 is intended to convey somewhat different information. Here we suppose that the practical interest is not in some measure of current risk but with the probability that a subject, having gone arrest-free for any given length of time, will now survive the whole of some pre-defined period after release. That period might be an individual's term on parole; but the illustration here takes it to be 3 years for all four subjects.

These results are statistical prognoses based on the experience of several thousand individuals. They could serve as a basis for policy guidelines, for example suggesting different "levels of intensity" of parole supervision to be normally assigned to different ranges of failure odds. Like medical prognoses, of course, they cannot predict the outcome of specific cases, leaving considerable discretion to the clinician -- in this case the parole officer.

We assume that most parole officers, given this kind of forecast information, would take it into consideration in coming to a set of decisions on how best to distribute their energies and available resources among the cases for which they are responsible. For example, one officer might adopt as a rule of thumb that his or her attention should be given simply in proportion to risk level. Another officer might decide on a triage practice, calculating that cases like 64 and 1235 are (for different reasons) likely to benefit least from his or her efforts and thus concluding that the greatest chance of making a difference lies with the statistically more uncertain middle risk group.

Presumably parole officers would also be interested in monitoring the results of their own tactics to determine whether in fact they are beating the odds with their case load. Such monitoring might be done simply by comparing the number of cases surviving arrest-free for one year (or any other fixed time of interest) with the number expected on the basis of the risk assessment. Somewhat more complicated, although perhaps more informative, might be a comparison of subjects' actual times to failure with the times expected on the basis of their assessed risk.

TABLE VI.1. CHARACTERISTICS OF SELECTED SUBJECTS

Variable	Case Number 1235	615	292	64
AGEFIRST	19.5	14.6	8.5	12.7
CARLNGTH	0.8	6.7	8.9	5.1
NOARRSTS	1	4	16	21
PPARVIOL	0	1	0	8
PRCOMMIT	0	1	2	0
Criminal History Scores:				
VIOLENCE	1	0.7	3	1
ROBBERY	0	1.4	0	0
BURGLARY	0	0	5.6	0
OTHPROP	0	0	4.6	2.7
GENDELQ	0	1	3.7	15.1
MWF	3	2	2	2
CYAVIOL	1	1	0	4
INFRRATE	0.8	0	2.2	0
TIMEIN	1.2	0.2	2.3	0.8
AGEOUT	21.4	21.5	19.6	18.6
ALCOHOL	2	2	0	2
DRUGS	2	2	1	2
GANG	0	1	0	2
DROPOUT	1	1	0	1
SCHDISC	0	2	0	2
FAMSIZE	0	1	1	1
FAMVIOL	2	0	0	0
PARALCH	1	0	0	1
PARCRIM	0	0	0	0
SIBCRIM	2	0	2	2
WEAKMOM	0	1	3	2
PCLRATE	0.16	0.19	0.14	0.22
PCRATE	58.9	65.8	65.3	49.4
VCLRATE	0.56	0.62	0.45	0.47
VCRATE	9.7	13.9	17.3	11.7
Region:	SONOTLA	SONOTLA	LA	NORCNTRL
Outcome:	No Arrest	Arrested	No Arrest	Arrested
Days After Rei.	1322	349	1547	314

Fig. VI.1
Six Month Ahead Forecast

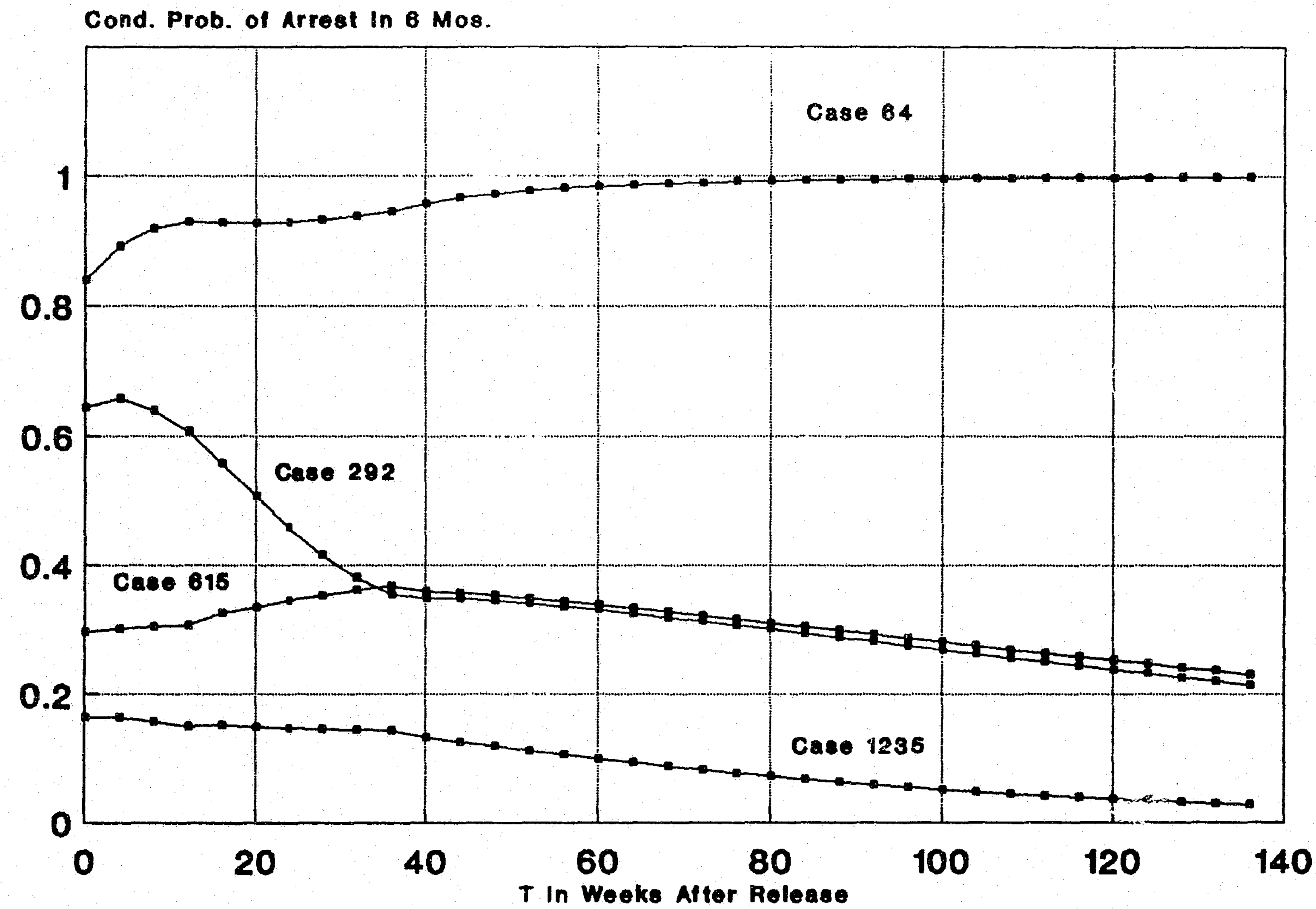
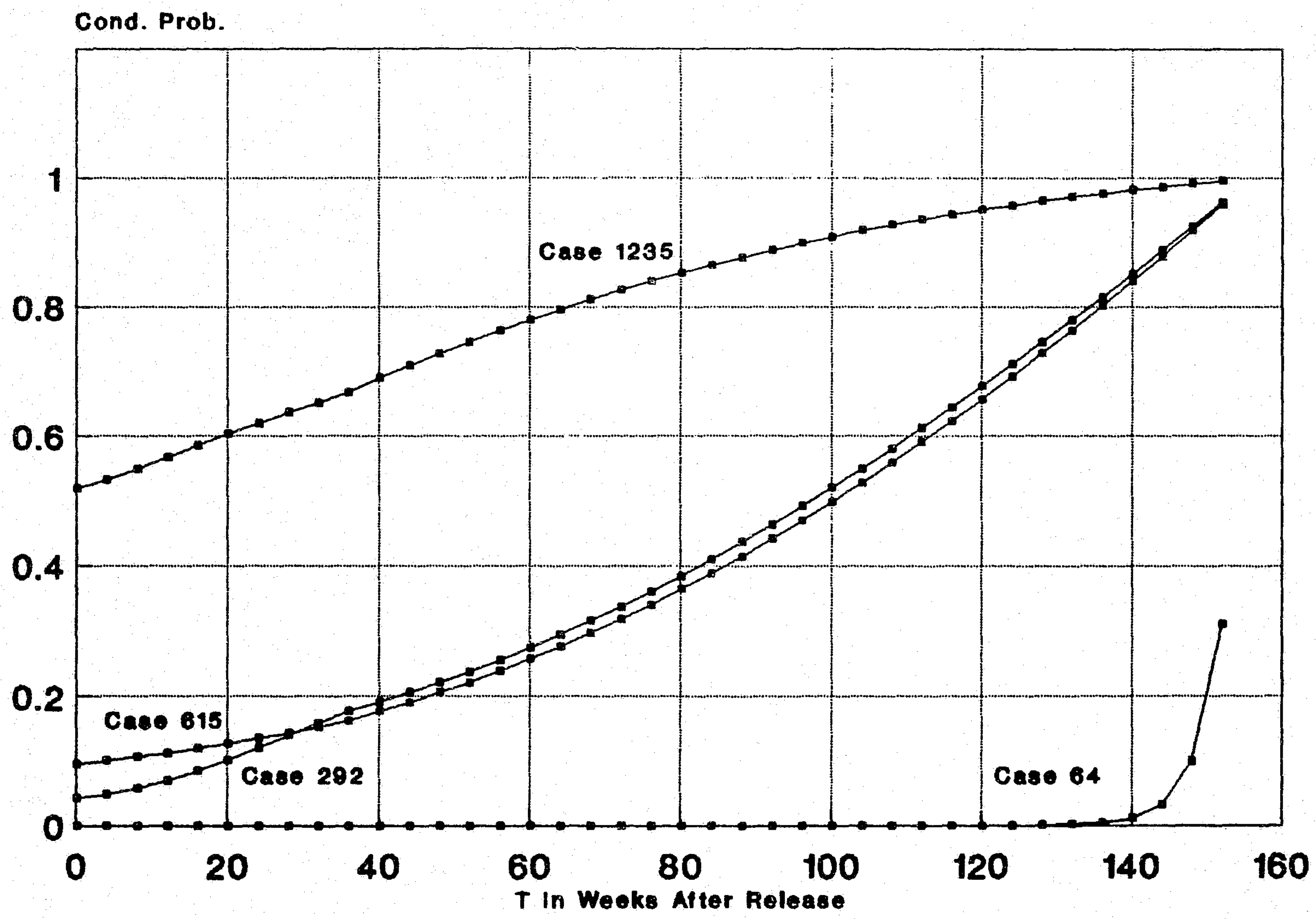


Fig. VI.2
156 Wk. Surv. Prob. Given Survival to T



For the administrator, two of the most persistent concerns are obtaining the staff and other resources needed and demonstrating that established policies in fact use those resources efficiently. As already remarked, the administrator may want to establish guidelines for the intensity and kind of supervision and services to be assigned individuals, based on an examination of the distribution of risk over the currently paroled population. But he or she might also use such model results as one basis for an examination of the case load distribution.

Still, the most important benefit to the administrator, we feel, lies in the potential a model such as this offers for economical policy "experimentation." Again we suppose the model adequately reflects current practice. If a change is instituted in the parole treatment of any reasonably sized, well-defined subset of the population, its effects (if any) ought to be detectable as significant deviations from expected failure rates within a relatively short time -- perhaps of the order of 4 to 6 months.[40]

As a simple example, suppose it were decided to test a policy under which supervision intensity would be markedly increased for a specified high risk group at the expense of reduced supervision of low risk subjects. The expected numbers of failures could be calculated separately for these two groups for each 4 week interval subsequent to implementation of the policy. If the new policy is indeed having a strong effect, differences between the observed and expected numbers of failures should become statistically significant within a relatively short time.

Such "experimentation," of course, is not always deliberate on the part of the administrator. Operational conditions change with changing characteristics or size of the population of wards released to parole. Sentencing practices may be altered drastically. Novel and only partially tested technologies for supervision and control are introduced -- sometimes in hope, but often in desperation. And perhaps the most universal concern: appropriated budgets do not always keep pace with increased work loads. Again we would suggest that statistical results derivable from a model like the one presented in this paper would allow the administrator to monitor the impact of such changes on recidivism and could form an important part of the basis for his demand for adequate resources to counter their effects.

It should be pointed out that these kinds of applications of a statistical prediction methodology require that it be based on a hazard function formulation. "Static" predictors (that is correlational models that do not explicitly take into account subjects' observed time to failure or censoring) may well be as accurate in assigning to individuals a probability of failure within some fixed period after release. But they cannot be adapted to answer questions such as "How many failures do we expect next month or over the next year?" without imposing some heroic assumptions that would essentially turn them into hazard models[41].

It was at one time argued quite convincingly that a correctional classification instrument should be reducible to a weighted, additive, paper-and-pencil scale that could be easily and quickly computed. With the ready availability of quite powerful personal computers this argument would seem to lose much of its force. Software could easily be developed so that in routine use individual case data would be entered and a pre-programed set of model-driven prognoses delivered automatically.

---

[40]Any such analysis, of course, would have to take care to assure that the outcome was not a "self-fulfilling prophecy." For example, if more intensive parole supervision was assigned to high-risk cases, the opportunity for detecting technical violations and, perhaps, new crimes as well would increase.

[41]Presumably, the reader will not be misled by the fact that in this paper failure is defined as the first rearrest after release. Models of this type can be developed in the same way for other definitions provided only that the event constituting a failure be unambiguously defined and locatable in time. Some examples might be the risk of a rearrest leading to a new conviction or sustained petition; or a rearrest but only on a felony charge; or perhaps a rearrest on a felony charge but only while the subject is still under parole supervision.

But as with any statistical predictor, it would be very important to maintain a realization among all users that such prognoses cannot see into the future. While developed for each subject on the basis of his characterization by the model's independent variables, they can "forecast" only what should be expected on average among a large number of similar cases.

6.2 Costs.

If the preceding paragraphs have suggested some of the benefits the authors see in the use of a hazard model in practise, what should be said about the costs?

Clearly a fairly substantial investment has to be made in assembling the large data base needed for estimating the model. Both the data to be used to describe case characteristics and the outcome data must be accurate and substantially complete. There may be a variety of outcomes (definitions of "failure") that would have a potential for useful application to policy issues. It would be important to ensure in advance that outcome data collected will satisfy all major demands for assessment of different kinds of risk.

Furthermore, the data should to the greatest extent possible reflect current conditions and policies. In this latter regard, the hazard formulation has a distinct advantage over "static" predictors. All subjects whose records are used in building the model need not have been at risk for the same length of time. The data base can include subjects who were released to parole fairly recently, thus giving some assurance that the population on which the model is built and the one to which its "predictions" are being applied are more nearly contemporaneous.

Of course, the model must also be re-estimated periodically -- perhaps every year or two. This task in itself involves no great expense but it implies that the data base, once assembled, must be maintained so that as a matter of routine it contains relevant individual history information on all new subjects entering the parole population and reasonably near-real-time information on failures.

Finally, resources must be made available to support a level of technical and research staff commensurate with the anticipated demands for analytic results that the model and the data base could provide.

It will undoubtedly come as no surprise to the reader that the authors think the benefits deriving from a substantially improved insight into the effectiveness of parole policies and practises could far outweigh these costs.

REFERENCES

Allison, P. D. (1984) Event History Analysis. Beverly Hills, Calif.: Sage Publications.

Baird, S. C., Storrs, G.M., and Connelly, H. (1984). Classification of Juveniles in Corrections: A Model Systems Approach. Final Report to the Office of Juvenile Justice and Delinquency Prevention. Washington, D.C.: Arthur D. Little, Inc.

Bishop, D.M., Frazier, C.E., and Henretta, J.C. (1989). "Prosecutorial waiver: case study of a questionable reform." Crime and Delinquency 35: 179-201.

Blumstein, A., Cohen, J., Roth, J. and Visher, C.A., eds. (1986). Criminal Careers and "Career Criminals". Volume I. Washington, D.C.: National Academy Press.

Box, G. and Jenkins, G. (1976). Time-Series Analysis. San Francisco, Calif.: Holden-Day.

Bureau of Justice Statistics (1987). Recidivism of Young Parolees. Special Report. Washington, D.C.: U.S. Department of Justice.

Bureau of Justice Statistics (1988). Report to the Nation on Crime and Justice. Second Edition. Washington, D.C.: U.S. Department of Justice.

Cook, P. and Laub, J. (1986). "The (surprising) stability of youth crime rates". Journal of Quantitative Criminology 2: 265-278.

Copas, J. (1985). "Prediction equations, statistical analysis and shrinkage". In David Farrington and Roger Tarling (eds.), Prediction in Criminology. Albany, N.Y.: State University of New York Press.

Copas, J. and Tarling, R. (1986). "Some methodological issues in making predictions." In Alfred Blumstein, Jacqueline Cohen, Jeffrey Roth, and Christy Visher (eds.), Criminal Careers and "Career Criminals", Vol. II. Washington, D.C.: National Academy Press.

Cronin, R.C., Bourque, B.B., Mill, J.M., Gragg, F.E., and McGrady, A.A. (1988). Evaluation of the Habitual Serious and Violent Juvenile Offender Program: Executive Summary. NCJ 105230. Washington, D.C.: Office of Juvenile Justice and Delinquency Prevention.

Eladt-Johnson, Regina C. and Johnson, Norman L. (1980). Survival Models and Data Analysis. New York: Wiley.

Elliott, D., Huizinga, D., and Ageton, S. (1985). Delinquency and Drug Use. Beverly Hills, Calif.: Sage Publications.

Farrington, David and Loeber, Rolf (1989) "Relative improvement over chance (RIOC) and phi as measures of predictive efficiency and strength of association in 2x2 tables." Journal of Quantitative Criminology 5: 201-214.

Farrington, David and Tarling, Roger eds. (1985). Prediction in Criminology. Albany, N.Y.: State University of New York Press.

Federal Bureau of Investigation (1989). Crime in the U.S., 1987. Washington, D.C.: U.S. Department of Justice.

Gottfredson, S. and Gottfredson, D. (1985). "Screening for risk among parolees: Policy, practice, and method." Chapter 2 in D. Farrington and R. Tarling (eds.), Prediction in Criminology. New York: Suny Press, Albany.

Gottfredson, S. and Gottfredson, D. (1986). "Accuracy of Prediction Models". In A. Blumstein, J. Cohen, J. Roth, and C. Visher (eds.), Criminal Careers and "Career Criminals", Vol. II. Washington, D.C.: National Academy Press.

Gottfredson, D., and Tonry, M. (1987). Prediction and Classification. Vol. 9 of Crime and Justice. Chicago, Ill.: University of Chicago Press.

Greenwood, P. (1986). "Predictors of chronic criminal behavior". Chapter 5 in P. Greenwood (ed.), Intervention Strategies for Chronic Juvenile Offenders. New York, N.Y.: Greenwood Press.

Guarino-Ghezzi, S., and Bryne, J.M. (1989). "Developing a model of structured decision making in juvenile corrections: The Massachusetts experience." Crime and Delinquency 35: 270-302.

Kalbfleisch, J.G. and Prentice, R.L. (1980). The Statistical Analysis of Failure Time Data. New York: John Wiley.

Kmenta, J. (1986). Elements of Econometrics, 2nd ed. New York: MacMillan.

Krisberg, B., Schwartz, I.M., Litsky, P., and Austin, J. (1986). "The watershed of juvenile justice reform." Crime and Delinquency 32: 5-38.

Larimore, W.E. (1983). "Predictive inference, sufficiency, entropy and an asymptotic likelihood principle". Biometrika
70: 175-181.

Larimore, W.E. and R.K. Mehra (1985). "The problem of overfitting data." Byte Oct.: 156-180.

Loeber, R. and T. Dishion (1983). "Early predictors of male delinquency: a review." Psychological Bulletin 94: 68-99.

Loeber, R. and M. Stauthaumer-Loeber (1986). "Family factors as correlates and predictors of juvenile conduct problems and delinquency." In M. Tonry and N. Morris (eds.), Crime and Justice, Vol. 7. Chicago, Ill.: University of Chicago Press.

London, Dick (1988). Survival Models and Their Estimation. Winsted, CN: Actex.

Maltz, M. (1984). Recidivism. Orlando, Fla.: Academic Press.

National Advisory Committee (NAC) for Juvenile Justice and Delinquency Prevention (1984). Serious Juvenile Crime: A Redirected Federal Effort. Washington, D.C.: Office of Juvenile Justice and Delinquency Prevention.

Nelson, James F. (1989). "An operational definition of prior criminal record." Journal of Quantitative Criminology 5(4):333-352.

Palmer, T. and Wedge, R. (1989). "California's juvenile probation camps: Findings and implications." Crime and Delinquency 35: 234-253.

Regnery, A.S. (1986). "A Federal perspective on juvenile justice reform." Crime and Delinquency 32: 39-51.

Schmidt, P. and Witte, A. (1988). Predicting Recidivism Using Survival Models. New York: Springer-Verlag.

Sechrest, L., White, S.O., and Brown, E., eds. (1979). The Rehabilitation of Criminal Offenders: Problems and Prospects. Washington, D.C.: National Academy Press.

Speirs, V.L. (1988). "Targeting serious juvenile offenders for prosecution can make a difference". NIJ Reports No. 211. Washington, D.C.: National Institute of Justice.

Visher, C.A., and Linster, R.L. (1990). "Survival model of pretrial failure." Journal of Quantitative Criminology 6:
153-184.

Weiner, N.A. (1989). "Violent criminal careers and 'violent career criminals'". Chapter 2 in Weiner, N. and Wolfgang, M., eds., Violent Crime, Violent Criminals. Newbury Park, Calif.: Sage Publications.

Wright, W.E., and Dixon, M.C. (1977). "Community treatment of juvenile delinquency: a review of evaluation studies." Journal of Research in Crime and Delinquency 19: 35-67.

APPENDIX A

A.1 The Likelihood Function.

By definition the likelihood function is the joint probability of occurrence assigned by the model to all observed outcomes. Let $S(t,Z)$ be the probability of survival to time t of a subject characterized by the covariate vector Z; $h(t,Z)\,dt$ the conditional probability of failure in the interval $[t, t+dt)$, given survival to time t; and $f(t,Z)\,dt$ the unconditioned probability of failure in $[t, t+dt)$. Then

$$f(t,Z) = h(t,Z) \times S(t,Z) \qquad \text{A.1}$$

Since

$$f(t,Z) = -\frac{dS(t,Z)}{dt} \qquad \text{A.2}$$

it follows that

$$S(t,Z) = e^{-\int_0^t h(x,Z)\,dx} \qquad \text{A.3}$$

Suppose subject i was rearrested at time t_i. According to the model the probability assigned to this failure event is $f(t_i,Z_i)\,dt$. In contrast suppose subject j had not yet been rearrested by the time the data were collected and that his time at risk was t_j. Under the model the probability of this occurring is $S(t_j,Z_j)$. Dropping the time interval factors dt, we can therefore write the likelihood function as

$$L = \prod_i^F f(t_i,Z_i) \times \prod_j^{N-F} S(t_j,Z_j) \qquad \text{A.4}$$

where the first product is over all subjects who were rearrested and the second over all successes[42]. Finally, if hS is substituted for f in the first product, the log of the likelihood becomes:

$$\ln L = \sum_i^F \ln h(t_i,Z_i) + \sum_j^N \ln S(t_j,Z_j) \qquad \text{A.5}$$

The first sum is over all failures and the second over the entire population.

In this paper we have used a hazard function $h(t,Z)$ whose form in the first 36 weeks at risk is different from its form in the later period. With

$$h(t,Z) = \begin{cases} h_1(t,Z) & t \le 36 weeks \\ h_2(t,Z) & t > 36 weeks \end{cases} \qquad \text{A.6}$$

the survival function becomes

[42]This assumes both that the individual outcomes are independent of one another and that the censoring mechanism is independent of subject risk. (See Kalbfleisch and Prentice, 1980.) Both assumptions seem reasonable in this application.

$$S(t,Z) = \begin{cases} e^{-\int_0^t h_1(x,Z)dx} & t \le 36\ weeks \\ e^{-\int_0^{36wks} h_1(x,Z)dx} \times e^{-\int_{36wks}^t h_2(x,Z)dx} & t > 36\ weeks \end{cases} \tag{A.7}$$

Note that the hazard function will in general have a discontinuity at 36 weeks. However, the survival function and all probabilities calculated over finite time intervals are continuous for all positive t.

When these functions are substituted into ln L, the log-likelihood function is decomposed into separate sums of terms involving only h_1 or h_2. Thus, ln L is maximized by separately maximizing the two functions

$$\ln L_1 = \sum_{i_1}^{F_1} \ln h_1(t_i,Z_i) - \sum_{j_1}^{N_1} \int_0^{t_j} h_1(x,Z_j)dx - \sum_{j_2}^{N_2} \int_0^{36\ wks} h_1(x,Z_j)dx \tag{A.8}$$

and

$$\ln L_2 = \sum_{i_2}^{F_2} \ln h_2(x,Z_i) - \sum_{j_2}^{N_2} \int_{36\ wks}^{t_j} h_2(x,Z_j)dx \tag{A.9}$$

In the first of these equation the sum over i_1 is the sum over all subjects rearrested within 36 weeks after release; the sum over j_1 is over subjects either failing or censored[43] during this period. The sum on j_2 is over all subjects surviving at least 36 weeks.

In the second equation the sum over i_2 is over those subjects who survived at least 36 weeks but were subsequently observed to fail; the second sum here is again over all subjects who were not rearrested in the first 36 weeks.

A.2 Analytic Form of the Hazard Function.

The basic functional form used for the hazard functions in this paper is

$$h(x) = e^{\gamma} x^{\alpha} e^{\beta x} \tag{A.10}$$

(See Visher and Linster, 1990 for an alternative application of this hazard model.) This form allows for considerable flexibility in the individual hazard functions that can be represented.

1. If α is less than zero, the function is very large at small values of x. Conversely, if α is greater than zero, the function increases from zero at x = 0 -- very rapidly if α is less than 1; slowly if α is greater than 1.

2. If β is less than zero, the function decreases to zero exponentially for very large x. It is exponentially increasing if β is greater than zero.

---

[43]In our data no subjects were censored in the early period so F_1 is equal to N_1.

3. If α and β have the same sign, the function is monotonic. If they are of opposite sign, it is U-shaped for $\alpha < 0$ and has a single maximum for $\alpha > 0$. In either case it passes through its extremum at $x = |\alpha|/|\beta|$.

4. If β is less than zero, h(x) can always be transformed into a function proportional to a gamma density. In that case its integral over all positive x converges to a finite limit:

$$\int_0^\infty h(x)dx = \frac{e^\gamma \times \Gamma(\alpha + 1)}{|\beta|^{(\alpha+1)}} \qquad \text{A.11}$$

Thus, the model does not dictate a priori that all subjects will eventually be rearrested.[44]

In the hazard function h_1 defining rearrest risk in the first 36 weeks after release, the variable x is equal to the time elapsed since release measured in years. In h_2

$$x = t - 0.6712 \qquad \text{A.12}$$

where t again is in years since release. This form is valid only for t > 0.6904 years = 36 weeks. The reason for this choice is that the function h(x) has a singularity at x = 0. Although it can "recover" from this very quickly, it seemed imprudent to force a discontinuity onto the hazard function at 36 weeks. Hence, h_2 was chosen to exclude 0 from its domain of definition.

In both h_1 and h_2 the parameters α, β, and γ are defined as linear functions of a vector of individual covariates Z:

$$\begin{aligned} \alpha &= Z'a \\ \beta &= Z'b \\ \gamma &= Z'c \end{aligned} \qquad \text{A.13}$$

Here **a**, **b**, and **c** are vectors of model coefficients. In order for the integral of h_1 to converge near t = 0, the vector of coefficients of the ln t term for this part of the model must be such that Z'**a** is greater than -1.

For both the early and later periods after release ln h(x) decomposes into a sum of terms, each of which depends only on a single covariate Z_k. The implication is that, if hypothetical subjects i and j differ only in their measures on covariate k,

$$\ln\left(\frac{h(Z_i,t)}{h(Z_j,t)}\right) = (Z_{ki} - Z_{kj}) \times (c_k + b_k t + a_k \ln t) \qquad \text{A.14}$$

This property is used as one measure of the relative importance the model attaches to covariate m in its assessment of risk.

A3. Parsimony in Model Identification

One of the problems encountered in the development of models with covariates is overfitting: modeling relationships found in a given data set that are simply random deviations attributable to

[44] In fact, the model used in the analyses of this paper associates with the "average" subject a finite probability of not being rearrested. See the discussion of the geometric mean hazard function in Chapter IV.

sampling from an idealized parent population. The procedure outlined below was followed as a partial empirical solution to this problem.[45]

1. The data base was randomly split into two non-overlapping sub-samples. Models were built separately on the two samples and cross-validated.

2. Let L(U,v) denote the value of the likelihood function on data sample U with coefficients **v** obtained by maximizing the log likelihood on data sample V. Let X and Y denote the two data samples. Each of the model's coefficients is in turn set equal to zero and approximations to the changes in the four functions ln L(X,**x**), ln L(X,**y**), ln L(Y,**x**) and ln L(Y,**y**) are calculated.

3. For any given coefficient its measure of "inconsistency" between relationships found in samples X and Y is defined to be the sum of these four first order changes. If the sum of all four changes is positive, the increase in the sum of cross-validation log likelihoods, ln L(X,**y**) and ln L(Y,**x**), on elimination of that parameter outweigh the sum of the decreases in ln L(X,**x**) and ln L(Y,**y**). The parameter is then considered as a candidate for elimination.

When any coefficient is set to zero, the magnitude of the decreases in ln L(X,**x**) and ln L(Y,**y**) increase monotonically with the amount of information lost by the elimination of that coefficient. Thus, for example, an **x** coefficient with a solid estimated t-statistic would in principle not be dropped from the model unless its elimination resulted in a substantial and more than compensating increase in ln L(Y,**x**). By the same token coefficients with similar but not quite equal values might be dropped if the loss of information is negligible for both construction models.

4. The parameter with the largest, positive sum of first order changes is eliminated and the process starts again, separately maximizing the likelihoods on X and Y with the reduced form of the model. This iteration continues until the sum of first order changes is negative for all remaining parameters.

The argument justifying the adoption of this procedure may be stated slightly differently. Since the log likelihood is a sum over observations,

$$\ln L(X,x) + \ln L(Y,x) = \ln L(X+Y,x) \qquad \text{A.15}$$

This function is proportional to the log of the probability of observing the outcomes on the combined data, given the perspective of the model that maximizes the likelihood on data set X alone. The function ln L(X+Y,**y**) is a similar probability on the same universe of observations, but from the perspective of the **y**-parameter model.

Now let $\mathbf{x}_{(k)}$ and $\mathbf{y}_{(k)}$ be the parameter vectors maximizing the likelihoods on X and Y respectively under the condition that the k^{th} parameter of the original set be constrained to be zero. We are interested in the functions: $\ln \{L(X+Y,\mathbf{x}_{(k)})/L(X+Y,\mathbf{x})\}$ and $\ln \{L(X+Y,\mathbf{y}_{(k)})/L(X+Y),\mathbf{y})\}$. A positive value for either of these functions indicates that the set of observed outcomes on the whole population is more probable with the reduced model than with the original. Since we have no reason for preferring results derived from either the X or Y data sets alone, the sum of these two functions is defined as the measure of "inconsistency" for the k^{th} parameter.

Each step in this iteration requires the separate calculation of sets of parameter vectors $\mathbf{x}_{(k)}$ and $\mathbf{y}_{(k)}$, where the index k ranges over all non-zero parameters remaining in the model. In principle, the components of each such vector are the likelihood maximizing values the remaining parameters would

[45]For a discussion of parsimony in model construction and the allied problem of shrinkage on validation, see Box and Jenkins (1976), Copas (1985) or Copas and Tarling (1986). For an information theoretic approach to the problem, see Larimore (1983) and Larimore and Mehra (1985).

assume after elimination of parameter k. To reduce the computational burden the changed parameter values used in these "inconsistency measures" are estimated with linearized equations: the first order Taylor series expansions of the vector of first derivatives of the respective likelihood functions.

Once all covariates were entered and a model was identified, consistent under this definition in the relationships found in the two construction samples, parameters values were calculated by maximizing the likelihood on the two samples combined. Coefficients with low estimated t statistics were then eliminated sequentially, using the likelihood ratio test at each step to test for significance of the loss of information in the reduced model. In the final form adopted for analysis, all parameters have asymptotic t-statistics with probabilities less than about 0.10.

When a covariate Z_k is first entered into the model, it has 3 parameters associated with it and the first test of consistency it must pass is usually the form of the time dependence of its contribution to the likelihood. For example, suppose that in one sample the maximizing value of the k^{th} coefficient of ln t (a_k) is negative but it is positive in the other. What this means is that sample 1's correlation of Z_k with high risk in the very early period after release is not found in sample 2 and is, therefore, suspect of being a peculiarity of sample 1 that is not generally found in the population. Perhaps a "better" model form would result if a_j were dropped. Indeed, most of parameters eliminated in the course of this identification process had opposite signs in the two construction samples.

As an illustration Table A.1 lists the terms eliminated in the first 8 models estimated for the earlier period at risk (h_1). At this point any further elimination of terms would in this approximation decrease the cross-validation likelihoods.

TABLE A.1. COEFFICIENTS ELIMINATED DURING MODEL DEVELOPMENT

Model Number	Coefficient	Data Set A Coefficient/(t-statistic)	Data Set B Coefficient/(t-statistic)
1	c: NOARRSTS	0.2658/(1.6039)	-0.0628/(1.0496)
2	a: NOARRSTS	0.0037/(0.2195)	-0.0246/(1.7833)
3	a: VIOLENCE	0.0814/(0.7762)	0.0449/(0.5477)
4	b: VIOLENCE	0.1086/(0.4331)	-0.3601/(1.8845)
5	b: PPARVIOL	0.1135/(0.3093)	0.2683/(0.7925)
6	c: ROBBERY	0.6435/(1.2385)	-0.6128/(2.1140)
7	b: GENDELQ	0.0797/(0.3358)	-0.3756/(1.7746)
8	a: ROBBERY	-0.0093/(0.2128)	0.0058/(0.9755)

By dropping these 8 terms from the 36 terms of the initial model (3 coefficients each for the 11 criminal history variables and the intercept term), the maximized log likelihood on data set A decreased from -410.73 to -413.43. For data set B the decrease was somewhat greater: from -401.07 to -408.38. Under a likelihood ratio test, there is no statistically significant loss of information from data set A (chi-squared probability = 0.71). The model on data set B, however, has apparently lost some of its global explanatory power (chi-squared probability = 0.07). The assumption underlying this process is, of course, that the detail that was lost was in both cases peculiar to that data set and not characteristic of the population from which the samples were drawn.

This process is considered only as a "semi-automated" guide to model identification. Even if the process were accepted in principle, the calculated results depend on the validity of a linear approximation to obtain the "test values" of a new set of coefficients. If the elimination of a particular coefficient would in fact produce considerable changes in the magnitudes of some of the other model coefficients, the truncation of the Taylor series with the linear term may be giving misleading results.

There were two kinds of cases in which the algorithm's decision to eliminate a term from the model was considered to be based on a dubious approximation:

1. A coefficient was eliminated whose estimated t-statistic indicated a reasonable level of significance in the model on one data set. In the other data set the coefficient was not close to being significant but was either of the same sign or had a value close to zero; and

2. The eliminated coefficient had low t-statistics but approximately equal values in the models built on the two data sets.

As the last step in model identification such terms were added back into the model built on the combined data. Of the 6 terms restored in the model for the early period at risk, all were subsequently re-eliminated because of low estimated t-statistics and lack of significance of the information added. In the model for the later period, 13 terms were restored. The only one reaching a reasonable level of significance was the coefficient of gang involement in the ln t term.

APPENDIX B
The Relative Improvement Over Chance (RIOC) Statistic

In a two way classification of N subjects with a given number E defined as the high risk group, the RIOC statistic is a measure of the improvement over random assignment to high and low risk classes normalized by the maximum improvement possible. Consider the 2x2 contingency table below.

Observed

		Fail	Survive	
Predicted	High Risk	A	B	E
	Low Risk	C	D	N-E
		F	N-F	N

We regard the E subjects classified as high risk to be "predicted" failures, the (N-E) low risk subjects to be "predicted" successes.

Suppose that, instead of using the model to define these classes, we simply picked E subjects at random from the population. On average we would expect that a fraction F/N would turn out to be failures so that by chance alone we would expect to make ExF/N correct failure predictions. The quantity (A - EF/N) is, therefore, the improvement in the number of correct predictions of failure that can be attributed to the model's risk classification.

The number A of "true positives" obviously cannot be larger than the <u>smaller</u> of the two numbers E (the number rated as high risk) or F (the number actually failing). Therefore, the greatest value the quantity (A - EF/N) can possibly attain is

$$\left(A_{max} - \frac{EF}{N}\right) = \begin{cases} \left(E - \frac{EF}{N}\right) & E < F \\ \left(F - \frac{EF}{N}\right) & F < E \end{cases} \qquad \text{B.1}$$

If A_{obs} is the observed number of high risk failures ("true positives"),

$$RIOC = \frac{\left(A_{obs} - \frac{EF}{N}\right)}{\left(A_{max} - \frac{EF}{N}\right)} \qquad \text{B.2}$$

Suppose the concern is with the number of "false positives" so that the number rated as high risk (E) is chosen to be less than the total number of failures (F). Then A_{max} = E and the RIOC becomes

$$RIOC = \frac{A - \frac{EF}{N}}{E\left(1 - \frac{F}{N}\right)} \qquad \text{B.3}$$

The denominator here is just the number of success one would expect among the E high risk subjects if the classification were purely random. Defining $\langle B \rangle$ to be the number of "false positives" expected by chance and substituting (E-B) for A, we can write

$$RIOC = \frac{(\langle B \rangle - B)}{\langle B \rangle} \qquad \text{B.4}$$

In this form the RIOC is interpretable directly as the fractional reduction the classification scheme produces in the number of "false positives."

If the classification scheme is designed to avoid "false negatives" so that the number of subjects rated as high risk is taken to be greater than the observed number of failures, we obtain in an exactly similar way

$$RIOC = \frac{(\langle C \rangle - C)}{\langle C \rangle} \qquad \text{B.5}$$

Since

$$\langle B \rangle = E\left(1 - \frac{F}{N}\right)$$

$$\langle C \rangle = F\left(1 - \frac{E}{N}\right) \quad and \qquad \text{B.6}$$

$$B = E - (F-C)$$

we obtain

$$\langle B \rangle - B = \langle C \rangle - C \qquad \text{B.7}$$

Thus, the expected decrease in the number of "false positives" is always matched by an identical decrease in the number of "false negatives" -- a result that is, perhaps, obvious.

Finally, using these results we can write for $E < F$

$$\frac{\langle C\rangle - C}{\langle C\rangle} = \frac{\langle B\rangle}{\langle C\rangle}\times RIOC$$

$$= \frac{E(N-F)}{F(N-E)}\times RIOC \qquad \text{B.8}$$

and for E > F

$$\frac{\langle B\rangle - B}{\langle B\rangle} = \frac{\langle C\rangle}{\langle B\rangle}\times RIOC$$

$$= \frac{F(N-E)}{E(N-F)}\times RIOC \qquad \text{B.9}$$

The coefficients of the RIOC on the right hand sides here are less than 1. Thus for E < F, the fractional reduction of false negatives is necessarily less than that of false positives and vice versa for E > F. The two are, of course, equal if E = F.